VISVESVARAYA TECHNOLOGICAL UNIVERSITY

Jnana Sangama, Belagavi – 590018, Karnataka, India

A PROJECT REPORT

On

"Sign Language Interpreter using Deep Learning"

A Dissertation Submitted in partial fulfilment of the requirement for the degree of

BACHELOR OF ENGINEERING

In

COMPUTER SCIENCE & ENGINEERING

Submitted by

Bahaduri Prachiti Jagdish (1RG17CS009)

Mohammed Azim (1RG17CS030)

Mohammed Touseef (1RG17CS031)

Prajwal B Mani (1RG17CS037)

Under The Guidance of

Mrs. Pushplata Dubey

Asst Professor, Dept of CSE

RGIT, Bengaluru-32

Department of Computer Science & Engineering

RAJIV GANDHI INSTITUTE OF TECHNOLOGY

Cholanagar, R.T. Nagar Post, Bengaluru-560032

2020-2021

RAJIV GANDHI INSTITUTE OF TECHNOLOGY

(Affiliated to Visvesvaraya Technological University)
Cholanagar, R.T. Nagar Post, Bengaluru-560032

Department of Computer Science & Engineering

CERTIFICATE

Phase- I

This is to certify that the Project Report titled **"Sign Language Interpreter using Deep Learning"** is a bonafide work carried out by **Ms. Bahaduri Prachiti Jagdish (USN 1RG17CS009), Mr. Mohammed Azim (USN 1RG17CS030), Mr. Mohammed Touseef (USN 1RG17CS031) and Mr. Prajwal B Mani (USN 1RG17CS037)** in partial fulfillment for the award of **Bachelor of Engineering** in **Computer Science and Engineering** under **Visvesvaraya Technological University, Belgavi**, during the year **2020-2021.** It is certified that all corrections/suggestions given for Internal Assessment have been incorporated in the report. This project report has been approved as it satisfies the academic requirements in respect of project work Phase- I (17CSP78) prescribed for the said degree.

Signature of Guide	Signature of HOD	Signature of Prinicpal
Mrs. Pushplata Dubey	**Mrs. Arudra A**	**Dr. Nagaraj A M**
Asst. Professor	Assoc. Professor & HOD	Principal
Dept. of CSE	Dept. of CSE	RGIT, Bengaluru
RGIT, Bengaluru	RGIT, Bengaluru	

Internal Evaluation

Name of the Examiners	**Signature with Date**
1.	
2.	

VISVESVARAYA TECHNOLOGICAL UNIVERSITY

Jnana Sangama, Belgavi 590018

RAJIV GANDHI INSTITUTE OF TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DECLARATION

We hereby declare that the project work entitled **"Sign Language Interpreter using Deep Learning"** submitted to the **Visvesvaraya Technological University, Belgavi** during the academic year **2020-2021**, is record of an original work done by us under the guidance of **Mrs. Pushplata Dubey,** Asst. Professor, Department of Computer Science and Engineering, RGIT, Bengaluru in the partial fulfilment of requirements for the award of the degree of **Bachelor of Engineering** in **Computer Science & Engineering.** The results embodied in this project have not been submitted to any other University or Institute for award of any degree or diploma.

Bahaduri Prachiti Jagdish (1RG17CS009)

Mohammed Azim (1RG17CS030)

Mohammed Touseef (1RG17CS031)

Prajwal B Mani (1RG17CS037)

ACKNOWLEDGEMENT

We take this opportunity to thank our college **Rajiv Gandhi Institute of Technology, Bengaluru** for providing us with an opportunity to carry out this project work.

We express our gratitude to **Dr. Nagaraj A M,** Principal, RGIT, Bengaluru and to **Dr. D G Anand**, Rector, RGIT, Bengaluru for providing the resources and support without which the completion of this project would have been a difficult task.

We extend our sincere thanks to **Mrs. Arudra A**, Associate Professor and Head, Department of Computer Science and Engineering, RGIT, Bengaluru, for being a pillar of support and encouraging us in the face of all adversities.

We would like to acknowledge the through guidance and support extended towards us by **Mrs. Pushplata Dubey**, Assistant Professor, Dept of CSE, RGIT, Bengaluru and project coordinators **Mrs. Geetha Pawar ,** Assistant Professor, Dept of CSE, RGIT, Bengaluru and **Mrs. Rajani Kodagali**, Assistant Professor, Dept of CSE, RGIT, Bengaluru. Their incessant encouragement and valuable technical support have been of immense help. Their guidance gave us the environment to enhance our knowledge and skills and to reach the pinnacle with sheer determination, dedication and hard work.

We also want to extend our thanks to the entire faculty and support staff of the Department of Computer Science and Engineering, RGIT, Bengaluru, who have encouraged us throughout the course of the Bachelor's Degree.

We want to thank our family for always being there with full support and for providing us with a safe haven to conduct and complete our project. We are ever grateful to them for helping us in these stressful times.

Lastly, we want to acknowledge all the helpful insights given to us by all our friends during the course of this project.

Bahaduri Prachiti Jagdish (1RG17CS009)
Mohammed Azim (1RG17CS030)
Mohammed Touseef (1RG17CS031)
Prajwal B Mani (1RG17CS037)

ABSTRACT

Speech Impairment is a disability, which affects an individual's ability to communicate using speech and hearing. This brings about the difficulty for both the sign and non - sign language speakers to communicate with each other. With recent advances in deep learning and computer vision, the focus of our project is to create a vision of an end to end Convolutional Neural Network that will be trained on the ASL(American Sign Language) dataset then modeled on robust architectures like GoogLeNet/MobileNet architecture and deploy it on an android application so that it will have more accessibility and provides an ease of use, thus aiding communication between signers and non-signers. It is a challenging and interesting problem that if solved will bring a leap in social and technological aspects alike.

CONTENTS

		Acknowledgement	**i**
		Abstract	**ii**
		List of Figures	**iv**
		List of Tables	**v**

CHAPTERS		TITLE	PAGE NO
1		**INTRODUCTION**	
1.1		**Introduction**	**1**
	1.1.1	**Deep Learning**	**2**
1.2		**Scope**	**2**
1.3		**Motivation**	**3**
1.4		**Problem Identification**	**3**
1.5		**Objectives**	**3**
	1.5.1	**Existing System**	**4**
	1.5.2	**Proposed System**	**5**
2		**OUTCOME OF THE PROJECT**	**7**
3		**LITERATURE SURVEY**	**8**
4		**SYSTEM REQUIREMENTS**	
4.1		**Hardware Requirements**	**20**
4.2		**Software Requirements**	**20**
4.3		**Hardware Requirements Specification**	**21**
4.4		**Software Requirements Specification**	**21**
5		**GANTT CHART**	**24**
		References	**25**

LIST OF FIGURES

SL NO.	PARTICULARS	PAGE NO.
Figure 1.1	Traditional – based Approach	4
Figure 1.2	Glove – based Approach	5
Figure 1.3	The Architecture of Sign Language Interpreter	5
Figure 1.4	Predict Pipeline Architecture	6
Figure 1.5	Speak Pipeline Architecture	6
Figure 5.1	Gantt Chart	24

LIST OF TABLES

SL NO.	PARTICULARS	PAGE NO.
Table 3.1	NumPy CNN Android : A Library for Straightforward Implementation of Convolutional Neural Networks for Android Devices	9
Table 3.2	Deep Learning for American Sign Language Fingerspelling Recognition System	10
Table 3.3	American Sign Language Video Hand Gestures Recognition using Deep Neural Networks	11
Table 3.4	American Sign Language Recognition using Deep Learning and Computer Vision	12
Table 3.5	American Sign Language Character Recognition using Convolutional Neural Network	13
Table 3.6	Optimization of Transfer Learning for Sign Language Recognition Targeting Mobile Platform	14
Table 3.7	A Deep Learning Based Video Classification System Using Multimodality Correlation Approach	15
Table 3.8	Hand Gesture Recognition Using Deep Learning	16
Table 3.9	Real-time American Sign Language Recognition with Convolutional Neural Networks	17
Table 3.10	Deaf Talk using 3D Animated Sign Language	18

INTRODUCTION

Chapter 1

INTRODUCTION

1.1 Introduction:

Interpreting is a translational activity in which one produces a first and final translation on the basis of a one-time exposure to an expression in a source language. The most common two modes of interpreting are simultaneous interpreting, which is done at the time of the exposure to the source language, and consecutive interpreting, which is done at breaks to this exposure. Language interpretation is defined by the International Standards Organization (ISO). It states that rendering a spoken or signed message into another spoken or signed language, preserving the register and meaning of the source language content. To do so we make use of ASL, which is American Sign Language. American Sign Language is a language that uses hand signs, facial expressions, and body postures to communicate ideas. American Sign Language not only connects the people to those who are deaf but also it serves as a membership card into a linguistic subculture of our society that not everyone is privileged to enjoy. The statistics presented by the International Standard Organization (ISO) show that over 5% of the world's population or 466 million people has disabling hearing loss. And it is expected to rise to 900 million by the year 2050 which is double as compared to the current statistics.

A sign language interpreter must accurately convey messages between two different languages. An interpreter is available for both hearing and deaf individuals. The act of interpreting occurs when a hearing person speaks, and an interpreter renders the speaker's meaning into sign language, or other forms used by the deaf party(ies). The interpreting also happens in reverse, when a deaf person signs, an interpreter renders the meaning expressed in the signs into the oral language for the hearing party, which is sometimes referred to as voice interpreting or voicing. The current statistics estimate that the actual people who are benefited from hearing aid, are only 17% of the total population while the remaining 83% of the people are suffering from the need and use of hea

1.1.1 Deep Learning

Deep learning (also known as deep structured learning) is part of a broader family of machine learning methods based on artificial neural networks with representation learning. Learning can be supervised, semi-supervised or unsupervised.

Deep-learning architectures such as deep neural networks, deep belief networks, recurrent neural networks and convolutional neural networks have been applied to fields including computer vision, machine vision, speech recognition, natural language processing, audio recognition, social network filtering, machine translation, bioinformatics, drug design, medical image analysis, material inspection and board game programs, where they have produced results comparable to and in some cases surpassing human expert performance.

Artificial neural networks (ANNs) were inspired by information processing and distributed communication nodes in biological systems. ANNs have various differences from biological brains. Specifically, neural networks tend to be static and symbolic, while the biological brain of most living organisms is dynamic (plastic) and analog.

1.2 Scope

ASL interpreters work in a large variety of environments, including medical, legal, educational, mental health, vocational, and other environments. Interpreting is often viewed as a practice profession which requires careful judgement of interpersonal and environmental factors as well as expertise in the skills of the profession itself. Degree programs in ASL Interpreting are available at colleges, universities, and technical schools across the country, ranging from associate degrees to master's degrees. In addition, interpreters work with mentors, attend workshops, and get certifications to become more adept, gain experience, and open additional career opportunities. Other interpreters are Children Of Deaf Adults (CODAs) and are typically exposed to ASL and Deaf culture at a young age, giving them an advantage over later learners. In recent years, much research has gone into discerning whether ASL interpreters have access to adequate, specialized, real-world training and career support systems to ensure success and protect against interpreter burnout. Many studies note interpreter reports of frustration with training that proved inadequate to deal with real-world problems, and a lack of professional support. To prevent interpreter burnouts, there are many alternatives ranging from the same human intervention process which costs 100 dollars per hour to interpreting hand gloves which cost 40,000 dollars per pair.

1.3 Motivation

Communication is one of the basic requirements for survival in society. Deaf and dumb people communicate among themselves using sign language but normal people find it difficult to understand their language. Extensive work has been done on American sign language recognition but Indian sign language differs significantly from American sign language.ISL uses two hands for communicating (20 out of 26) whereas ASL uses single hand for communicating. Using both hands often leads to obscurity of features due to overlapping of hands. In addition to this, lack of datasets along with variance in sign language with locality has resulted in restrained efforts in ISL gesture detection. Our project aims at taking the basic step in bridging the communication gap between normal people and deaf and dumb people using Indian sign language. Effective extension of this project to words and common expressions may not only make the deaf and dumb people communicate faster and easier with outer world, but also provide a boost in developing autonomous systems for understanding and aiding them.

1.4 Problem Identification

According to the projection of data it is expected to rise to 900 million by 2050 that is double the current stats. Over 5% of the world's population which is almost 466 million people has disabling hearing loss. Studies reveal that deaf people are around twice as likely to suffer from psychological problems such as depression and anxiety. Hearing loss can affect a person in three main ways firstly, fewer educational and job opportunities due to impaired communication. Social withdrawal due to reduced access to services and difficulties communicating with others emotional problems caused by a drop in self-esteem and confidence.

1.5 Objectives

1. Create a robust android application that aids in communication for deaf people.
2. Use deep learning with transfer learning techniques to build a neural network model that as to ability to learn pattern in video and classify the images.
3. An error correction model for identifying the pattern mismatch and correcting it for the audio or text input format.
4. To build a model that has higher accuracy with less bias and overfitting problems.
5. To enhance the model to classify at least 20 classes or more.

1.5.1 Existing systems

There has been much research on the sign language domain. Some of them are traditional ones while some of them are advanced mechanics like sensor gloves. Therefore, these conventional mechanisms cannot effectively deal with the ever-evolving technologies.

1. **TRADITIONAL – BASED APPROACH**

   A sign language interpreter is a person trained in translating between a spoken and a signed language. This usually means somone who interprets what is being said and signs it for someone who can't hear, but understands the sign. The interpreter of course will also interpret and speak the words which convey the meaning of whatever the signing person signs, so that the hearing person can "hear" what is being signed.

2. **GLOVE – BASED APPROACH**

   The signers are required to wear a sensor glove or a coloured glove. The task will be simplified during segmentation process by wearing glove.

Disadvantages

- Lack of availability of the interpreter for that moment.
- The gloves based approach is that the signer has to wear the sensor hardware along with the glove during the operation of the system.
- The cost of traditional based approach will vary from 100 to 200 dollars/hrs and for gloves based approach its costs around 40,000 dollars / pair.

Fig. 1.1: Traditional – based Approach

Fig. 1.2: Glove – based Approach

1.5.2 Proposed system

We have introduced a refined approach using advanced deep learning techniques that is flexible to use by any mobile users. In this presentation, we propose an architecture to detect signs. Image pre-processing makes the existing data and input data normalized. The pipeline allows us to automate machine learning workflow. Transfer learnings make it easy to make the models to learn even small details that are hard to capture in a small network. Kivy gives us an advantages overview by giving us a docile structure to compile to any version of the mobile app.

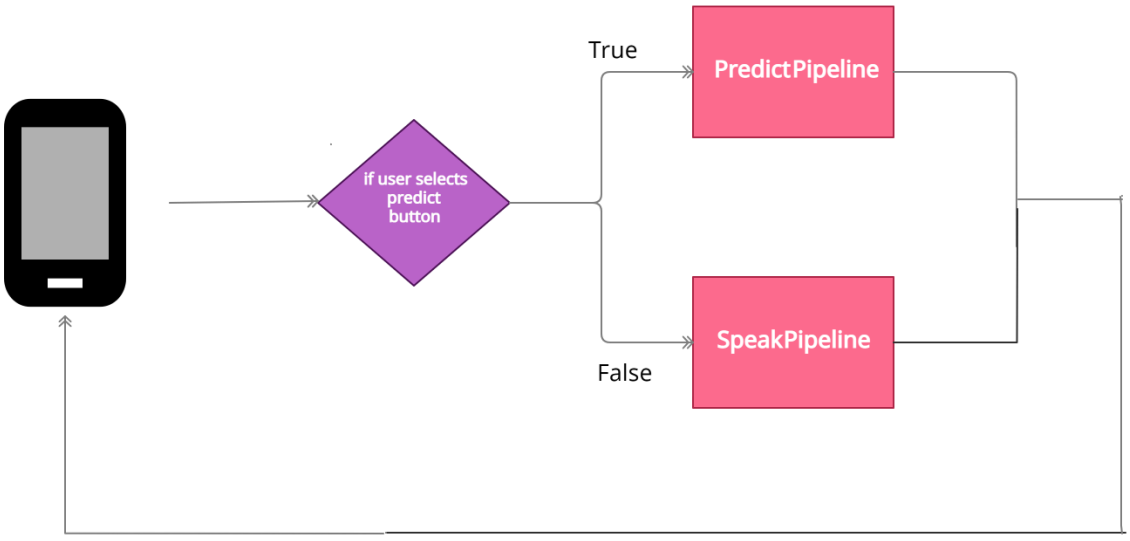

Fig. 1.3: The Architecture of Sign Language Interpreter

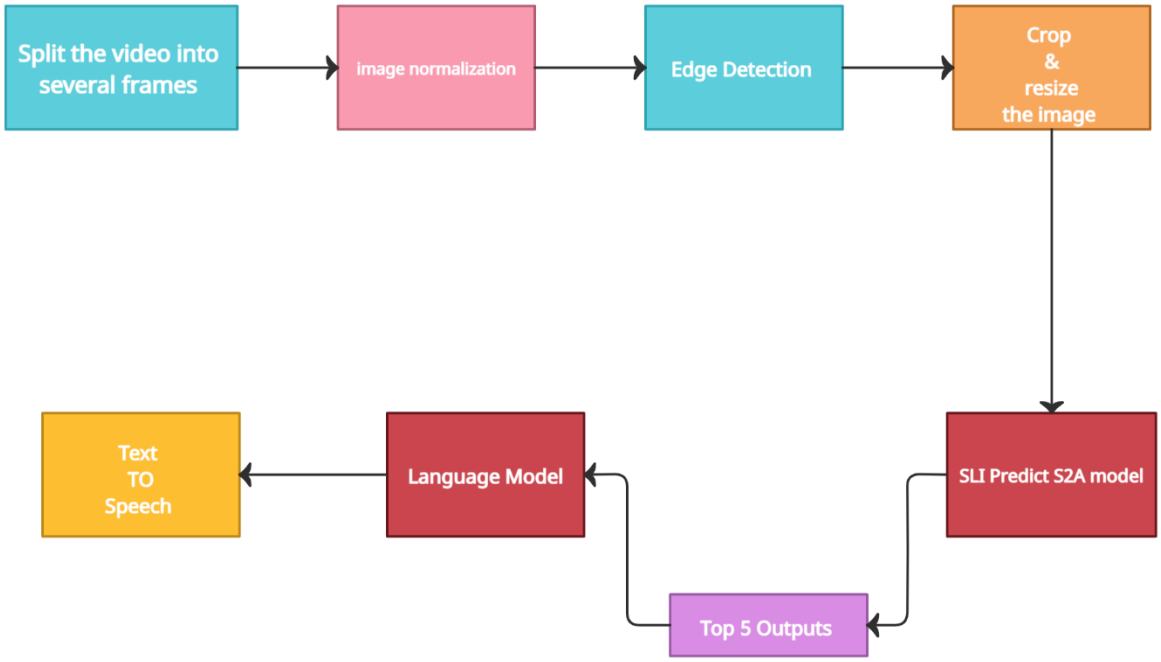

Fig. 1.4: Predict Pipeline Architecture

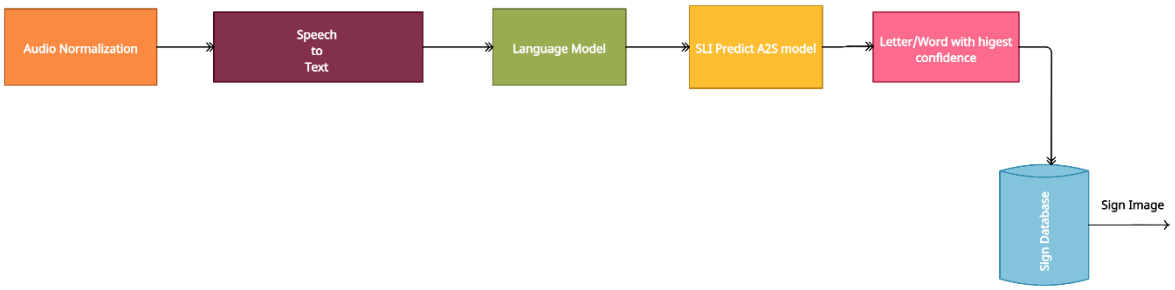

Fig. 1.5: Speak Pipeline Architecture

Advantages

- Considerably better accuracy compared to the previous models.
- Considerably less requirement of raw computation and Storage space requirements.
- Usage of Neural network model leads to less human intervention.

OUTCOME
OF PROJECT

Chapter 2

OUTCOME OF THE PROJECT

Sign Language Interpreter will be able to predict the sign waved by the user and converts that into sound or text vice versa. So far the systems that have been developed give less accuracy and slower results with only shallow networks. The model that has been used in our project will be able to detect with improved accuracy. We will achieve the following results:

1. **UI Interface to access model functionalities:** A flexible UI with a good look and feel with a design that is self-explanatory.
2. **Improved accuracy:** By building a deeper network that model learns and minimizes more parameter that is optimal for our problem
3. **Dynamic Tracing:** By allowing users to dynamical trace would add an advantage when the app is deployed in the real-time
4. **Continuous integration pipelines that enable model redeployment:** Pipelines helps us to automate the machine learning workflows which helps us to reduce time on the surface level.

LITERATURE SURVEY

Chapter 3

LITERATURE SURVEY

A literature survey or a literature review in a project report is that section which shows the various analyses and research made in the field of interest and the results already published, taking into account the various parameters of the project and the extent of the project. It is the most important part of a report as it gives a direction in the area of research. It helps to set a goal for analysis thus giving a problem statement. A literature review is both a summary and explanation of the complete and current state of knowledge on a limited topic as found in academic books and journal articles. Literature survey is a text of a scholarly paper, which includes the current knowledge including substantive findings, as well as theoretical and methodological contributions to a particular topic. And in general, a literature survey guides or helps the researcher to define/find out/identify a problem. The purpose of a literature survey is to: Place each work in the context of its contribution to understanding the research problem being studied. Describes the relationship of each work to the others under consideration. Identify new ways to interpret prior research. Literature of knowledge has a function to teach. It means that literature gives particular values, messages, and themes to the readers. The final function of literature is that literature relieve human either writers or readers from the pressure of emotions. Literature also functions to contribute values of human lives. It provides an excellent starting point for researchers beginning to do research in a new area by forcing them to summarize, evaluate, and compare original research.

As the technology progresses, the internet is now commonly used on PCs, tablets, and smartphones. This generates a huge amount of data, especially textual data. It has become impossible to manually analyze all the data for a specific purpose. New research directions have emerged from automatic data analysis like automatic emotion analysis. Emotion analysis has attracted researchers' attention because of its applications in different fields. Components such as behavior, voice, posture, vocal intensity, and emotion intensity of the a person depicting the emotion, when combined, helps in measuring and recognizing various emotions.

NumPy CNN Android : A Library for Straightforward Implementation of Convolutional Neural Networks for Android Devices

Table 3.1: NumPy CNN Android : A Library for Straightforward Implementation of Convolutional Neural Networks for Android Devices

Name of the paper	Authors	Year	Features	Advantages	Disadvantages
NumPy CNN Android : A Library for Straightforward Implementation of Convolutional Neural Networks for Android Devices	Ahmed Fawzy Gad	2019	AlexNet , VGGNet a nd GoogLeNet are examples of CNN architecture s trained with the ImageNet dataset.	The process of creating a CNN model working on mobile devices is straight forward as the trained model is compatible with them without any in-between transformation steps. Mean error across the samples used is 0.5113.	The weakness of the proposed library is its computational time.

Authors : Ahmed Fawzy Gad

A new open source library called NumPyCNNAndroid is proposed that minimizes the overhead of building and running convolutional neural networks on Android devices. The library is written in Python 3. It uses Kivy for building the application interface and Numerical Python for building the network itself. The library supports the most common layers. Compared to the widely known deep learning libraries, NumPyCNNAndroid avoids the extra overhead of making the network suitable for running on mobile devices. The experimental results validate the correctness of the library implementation by comparing results from both the proposed library and TensorFlow based on mean absolute error.

Deep Learning for American Sign Language Fingerspelling Recognition System

Table 3.2: Deep Learning for American Sign Language Fingerspelling Recognition System

Name of the paper	Authors	Year	Features	Advantages	Disadvantages
Deep Learning for American Sign Language Fingerspelling Recognition System	Huy B.D Nguyen, Hung Ngoc Do	2019	Histograms of Oriented Gradients(HOG) and Local Binary Pattern(LBP), Moore Neighbourhood algorithm, SIFT algorithm	The combinations of HOG, LBP features and multi-kernel multi-class VM acts as a good standalone Feature extractor, better result can be obtained than using an end-to-end CNN architecture.	The validation accuracy of the CNN-SVM model is lower than that of HOG-LBP-SVM model, it has a better chance to counter overfitting.

Authors - Huy B.D Nguyen, Hung Ngoc Do

Sign language has always been a major tool for communication among people with disabilities. In this paper, a sign language fingerspelling alphabet identification system would be developed by using image processing technique, supervised machine learning and deep learning. In particular, 24 alphabetical symbols are presented by several combinations of static gestures (excluding 2 motion gestures J and Z). Histogram of Oriented Gradients (HOG) and Local Binary Pattern (LBP) features of each gesture will be extracted from training images. Then Multiclass Support Vector Machines (SVMs) will be applied to train these extracted data. Also, an end-to-end Convolutional Neural Network (CNN) architecture will be applied to the training dataset for comparison. After that, a further combination of CNN as feature descriptor and SVM produces an acceptable result. The Massey Dataset is implemented in the training and testing phases of the whole system.

American Sign Language Video Hand Gestures Recognition using Deep Neural Networks

Table 3.3: American Sign Language Video Hand Gestures Recognition using Deep Neural Networks

Name of the paper	Authors	Year	Features	Advantages	Disadvantages
American Sign Language Video Hand Gestures Recognition using Deep Neural Networks	Shivashankara S & Srinath S	2019	Speeded Up Robust Features(SURF), Zernike Moment(ZM), Discrete Cosine Transform(DCF), Radon Features	Due to advance preprocessing the PVGR has an average accuracy of 96.43 Overall, the recognition rate obtained is better comparing with the state of art techniques.	It is noticed that, due to the gestures captured in low illumination night time, there is a bit of loss of recognition rate.

Authors : Shivashankara S & Srinath S

In this paper an effort has been placed to translate / recognize some of the video based hand gestures of American Sign Language (ASL) into human and / or machine readable English text using deep neural networks. Initially, the recognition process is carried out by fetching the input video gestures. In the recognition process of the proposed algorithm, for background elimination and foreground detection, the Gaussian Mixture Model (GMM) is used. The basic preprocessing operations are used for better segmentation of the video gestures. The various feature extraction techniques like, Speeded Up Robust Features (SURF), Zernike Moment (ZM), Discrete Cosine Transform (DCT), Radon Features (RF), and R, G, B levels are used to extract the hand features from frames of the video gestures. The extracted video hand gesture features are used for classification and recognition process in forthcoming stage. For classification and followed by recognition, the Deep Neural Networks (stacked autoencoder) is used. This video hand gesture recognition system can be used as tool for filling the communication gap between the normal and hearing impaired people. As a result of this proposed ASL video hand gesture recognition (VHGR), an average recognition rate of 96.43% is achieved. This is the better and motivational performance compared to state of art techniques.

American Sign Language Recognition using Deep Learning and Computer Vision

Table 3.4: American Sign Language Recognition using Deep Learning and Computer Vision

Name of the paper	Authors	Year	Features	Advantages	Disadvantages
American Sign Language Recognition using Deep Learning and Computer Vision	Kshitij Bantupalli & Ying Xie	2018	Use of Inception net, a CNN for recognizing spatial features. then use a RNN to train on temporal features. The dataset used is the American Sign Language Dataset.	Softmax layer gives more accuracy than the pool layer	One of the problems the model faced is with facial features and skin tones. While testing with different skin tones, the model dropped accuracy if it hadn't been trained on a certain skin tone and was made to predict it.

Authors : Kshitij Bantupalli & Ying Xie

Speech impairment is a disability which affects an individuals ability to communicate using speech and hearing. People who are affected by this use other media of communication such as sign language. Although sign language is ubiquitous in recent times, there remains a challenge for non-sign language speakers to communicate with sign language speakers or signers. With recent advances in deep learning and computer vision there has been promising progress in the fields of motion and gesture recognition using deep learning and computer vision-based techniques. The focus of this work is to create a vision- based application which offers sign language translation to text thus aiding communication between signers and non-signers. The proposed model takes video sequences and extracts temporal and spatial features from them. We then use Inception, a CNN (Convolutional Neural Network) for recognizing spatial features. We then use a RNN (Recurrent Neural Network) to train on temporal features. The dataset used is the American Sign Language Dataset.

American Sign Language Character Recognition using Convolutional Neural Network

Table 3.5: American Sign Language Character Recognition using Convolutional Neural Network

Name of the paper	Authors	Year	Features	Advantages	Disadvantages
American Sign Language Character Recognition using Convolutional Neural Network	Sarfaraz Masood, Manish Chandra Thuwal, Adhyan Srivastava	2018	Convolutional Neural Network Model, Image Preprocessing and VGG16 Model	The Convolution Neural Network provides a remarkable accuracy in identifying the sign language characters including alphabets and numerals.	Even though the accuracy is high but the percentage of misclassification is more than expected

Authors : Sarfaraz Masood, Manish Chandra Thuwal, Adhyan Srivastava

Inability to speak is considered to be a true disability. People with this disability use different modes to communicate with others, there are number of methods available for their communication one such common method of communication is sign language. Developing sign language application for deaf people can be very important, as they‟ll be able to communicate easily with even those who don‟t understand sign language. This work aims at taking the basic step in bridging the communication gap between normal people and deaf and dumb people using sign language. The image dataset consists of 2524 ASL gestures. The accuracy of the model obtained using Convolution Neural Network was 96%.

Optimization of Transfer Learning for Sign Language Recognition Targeting Mobile Platform

Table 3.6: Optimization of Transfer Learning for Sign Language Recognition Targeting Mobile Platform

Name of the paper	Authors	Year	Features	Advantages	Disadvantages
Optimization of Transfer Learning for Sign Language Recognition Targeting Mobile Platform	Dhruv Rathi	2018	Convolutional Neural Network(CNN) on ImageNet, two pre-trained models, Inception V3 model and MobileNets open-source models for comparison purposes.	Both mobile net and inception gives us an accuracy more than 95% irrespective of there parameters difference	Model can't predict sentences due to lack of Natural Language Processing(NLP) and other sentences construction techniques.

Authors : Dhruv Rathi

The target of this research is to experiment, iterate and recommend a system that is successful in recognition of American Sign Language(ASL). It is a challenging as well as an interesting problem that if solved will bring a leap in social and technological aspects alike. In this paper, we propose a real-time recognizer of ASL based on a mobile platform, so that it will have more accessibility and provides an ease of use. The technique implemented is Transfer Learning of new data of Hand gestures for alphabets in ASL to be modelled on various pre-trained high- end models and optimize the best model to run on a mobile platform considering the various limitations of the same during optimization. The data used consists of 27,455 images of 24 alphabets of ASL. The optimized model when ran over a memory-efficient mobile application, provides an accuracy of 95.03% of accurate recognition with an average recognition time of 2.42 seconds. This method ensures considerable discrimination in accuracy and recognition time than the previous research.

A Deep Learning Based Video Classification System Using Multimodality Correlation Approach

Table 3.7: A Deep Learning Based Video Classification System Using Multimodality Correlation Approach

Name of the paper	Authors	Year	Features	Advantages	Disadvantages
A Deep Learning Based Video Classification System Using Multimodality Correlation Approach	Jungheon Lee, Youngsan Koh and Jihoon Yang	2017	Image feature vector extraction, Audio feature vector extraction, Normalization of feature vectors	Method of normalizing each modality to a unit vector so that it can be effectively integrated at the integration stage of the multimodality.	Unit vector normalization should be done to get the required accuracy.

Authors : Jungheon Lee, Youngsan Koh and Jihoon Yang

In this paper, we propose a video event classification system that classifies video events using the correlation of images and sounds extracted from one video. The proposed system has better classification performance than other systems using single modality.

Hand Gesture Recognition Using Deep Learning

Table 3.8: Hand Gesture Recognition Using Deep Learning

Name of the paper	Authors	Year	Features	Advantages	Disadvantages
Hand Gesture Recognition Using Deep Learning	Soeb Hussain and Rupal Saxena	2017	Transfer Learning : VGG16 architecture	The accuracy is far better than video based standard network alexnet by scoring 93.09%	The method was made robust by avoiding skin color segmentation, blob detection, skin area cropping and centroid extraction for unidirectional dynamic gestures.

Authors : Soeb Hussain and Rupal Saxena

In order to offer new possibilities to interact with machine and to design more natural and more intuitive interactions with computing machines, our research aims at the automatic interpretation of gestures based on computer vision. In this paper, we propose a technique which commands computer using six static and eight dynamic hand gestures. The three main steps are: hand shape recognition, tracing of detected hand (if dynamic), and converting the data into the required command. Experiments show 93.09% accuracy.

Real-time American Sign Language Recognition with Convolutional Neural Networks

Table 3.9: Real-time American Sign Language Recognition with Convolutional Neural Networks

Name of the paper	Authors	Year	Features	Advantages	Disadvantages
Real-time American Sign Language Recognition with Convolutional Neural Networks	Brandon Garcia & Sigberto Alarcon Viesca (Stanford University Stanford, CA)	2016	A deep learning framework, in order to develop, test, and run our CNN specifically, they used Berkeley Vision and Learning Center's GoogLeNet pre-trained on the 2012 ILSVRC dataset.	Produce effective accuracy with the letters a-e without any image pre-processing	The lack of variation in our datasets, the validation accuracies we observed during training were not directly reproducible upon testing on the web application.

Authors - Brandon Garcia & Sigberto Alarcon Viesca (Stanford University Stanford, CA)

A real-time sign language translator is an important milestone in facilitating communication between the deaf community and the general public. We hereby present the development and implementation of an American Sign Language (ASL) fingerspelling translator based on a convolutional neural network. We utilize a pre-trained GoogLeNet architecture trained on the ILSVRC2012 dataset, as well as the Surrey University and Massey University ASL datasets in order to apply transfer learning this task. We produced a robust model that consistently classifies letters a-e correctly with first-time users and another that correctly classifies letters a-k in a majority of cases. Given the limitations of the datasets and the encouraging results achieved, we are confident that with further research and more data, we can produce a fully generalizable translator for all ASL letters.

Deaf Talk using 3D Animated Sign Language

Table 3.10: Deaf Talk using 3D Animated Sign Language

Name of the paper	Authors	Year	Features	Advantages	Disadvantages
Deaf Talk using 3D Animated Sign Language	Mateen Ahmed, Mujtaba Idrees, Zain ul Abideen, Rafia Mumtaz, Sana Khalique	2016	With Microsoft's Kinect in the market, especially for hands,used the depth sensor of Kinect to recognize around thousand phrases from ASL and this recognition is based on the hidden Markov model (HMM),also focused on Kinect base hand gestures recognition.	The system provides dual mode of communication so it has been categorized into two independent modules with 84% accuracy	This does not solve the communication problem, as natural language speakers don't understand sign language hence there exists a communication gap between these two communities.

Authors : Mateen Ahmed, Mujtaba Idrees, Zain ul Abideen, Rafia Mumtaz, Sana Khalique

This paper describes a neoteric approach to bridge the communication gap between deaf people and normal human beings. In any community there exists such group of disable people who face severe difficulties in communication due to their speech and hearing impairments. Such people use various gestures and symbols to talk and receive their messages and this mode of communication is called sign language. Yet the communication problem doesn't end here, as natural language speakers don't understand sign language resulting in a communication gap. Towards such ends there is a need to develop a system which can act as an interpreter for sign language speakers and a translator for natural language speakers. For this purpose, a software based solution has been developed in this research by exploiting the latest technologies from Microsoft i.e. Kinect for windows V2. The proposed system is dubbed as Deaf Talk, and it acts as a sign language interpreter and translator to provide a dual mode of communication between sign language speakers and natural language speakers. The dual mode of communication has following independent modules (1) Sign/Gesture to speech conversion (2) Speech to sign language conversion. In sign to speech conversion module, the person with speech inhibition has to place himself within Kinect's field of view (FOV) and then performs the sign language gestures. The system receives the performed gestures through Kinect sensor and then comprehends those gestures by comparing them with the trained

gestures already stored in the database. Once the gesture is determined, it is mapped to the keyword corresponding to that gesture. The keywords are then sent to text to speech conversion module, which speaks or plays the sentence for natural language speaker. In contrast to sign to speech conversion, the speech to sign language conversion module translates the spoken language to sign language. In this case, the normal person places himself in the Kinect sensor's FOV and speaks in his native language (English for this case). The system then converts it into text using speech to text API. The keywords are then mapped to their corresponding pre-stored animated gestures and then animations are played on the screen for the spoken sentence. In this way the disable person can visualize the spoken sentence, translated into a 3D animated sign language. The accuracy of Deaf Talk is 87 percent for speech to sign language conversion and 84 percent for sign language to speech conversion.

SYSTEM REQUIREMENTS

Chapter 4

SYSTEM REQUIREMENTS

To be used efficiently, all computer software needs certain hardware components or other software resources to be present on a computer these prerequisites are known as system requirements. System requirements are the configuration that a system must have in order for a hardware or software application to run smoothly and efficiently. Failure to meet these requirements can result in installation problems or performance problems. The former may prevent a device or application from getting installed, whereas the latter may cause a product to malfunction or perform below expectation or even to hang or crash. System requirements are also known as minimum system requirements. System requirements only tell what system must have and what it must allow users to do. The system requirements are of two types:

- Hardware Requirements
- Software Requirements

4.1 Hardware Requirements

The most common set of requirements defined by any operating system or software application is the physical computer resources, also known as hardware.

For Typical Operating System (Windows 10)

- Intel Core i3 3rd gen processor or later.
- 8GB RAM & 500GB disk space
- A good GPU support
- I/O devices
- Any external or inbuilt camera with minimum pixel resolution 200 x 200 (300pi or 1501pi) 4-megapixel cameras and up.

4.2 Software Requirements:

The Software requirements deal with defining software resource requirements and prerequisites that need to be installed on a computer to provide optimal functioning of an application.

- Front End : Android (using Kivy)
- Video & image processing : Open CV, scikit-image

- Neural Network Modeling : Tensorflow(tf.lite), Keras, Pytorch
- Architectures : Google LeNet, MobileNet, VGG 19
- Ubuntu : deploying .apk using buildozer kivy
- Editor : Visual code, jupyter notebook

4.3 Hardware Requirement Specification

Above mentioned requirements are minimum. Practically I shall recommend the following:

1. Minimum 2-GHz processor
2. RAM 4 GB
3. Hard disk available space minimum of 100 GB

- **Intel® Core™ Processors - The Powerful Processor:**
  Intel Core processors first came to the desktop in mid-2006, replacing the Pentium line that had previously comprised Intel's high-end processors. The Core "i" names are primarily "high level" categorizations that help differentiate processors within a given generation.
- **GPU:** A graphics processing unit (GPU) is a specialized, electronic circuit designed to rapidly manipulate and alter memory to accelerate the creation of images in a frame buffer intended for output to a display device. GPU will help us to train the models faster without using CPU resources.
- **I/O devices**: To view and input data to the application.
- **RAM**: RAM,also known as random access memory is vital for any system to function.
- **Minimum RAM requirements for our system**: 8 GB
- **Camera**: Most importantly you need a camera that can do 30 fps - that means 30 frames per second. Any less than 24 frames per second and your hands will be blurry.
- **HardDisk:** Minimum requirements for disk capacity is 500GB.

4.4 Software Requirement Specification

- **Front End : Android (using Kivy):** When you create an application with Kivy, you're creating a Natural User Interface or NUI. The idea behind a Natural User Interface is that the user can easily learn how to use your software with little to no instruction.
- **Ubuntu : deploying .apk using buildozer kivy:** You can create a package for android using the python-for-android project. This page explains how to download and use it directly on your own machine (see Packaging with python-for-android) or use the

Buildozer tool to automate the entire process. You can also see Packaging your application for the Kivy Launcher to run kivy programs without compiling them. Kivy applications can be released on an Android market such as the Play store, with a few extra steps to create a fully signed APK. The Kivy project includes tools for accessing Android APIs to accomplish vibration, sensor access, texting etc. These, along with information on debugging on the device, are documented at the main Android page.

- **Buildozer:** It is a tool that automates the entire build process. It downloads and sets up all the prerequisites for python-for-android, including the android SDK and NDK, then builds an apk that can be automatically pushed to the device. Buildozer currently works only in Linux, and is an alpha release, but it already works well and can significantly simplify the apk build.
- **Video & image processing: Open CV, scikit-image:** Processing a video means, performing operations on the video frame by frame. Frames are nothing but just the particular instance of the video in a single point of time. We may have multiple frames even in a single second. Frames can be treated as similar to an image

1. Adaptive Threshold

By using this technique we can apply thresholding on small regions of the frame. So the collective value will be different for the whole frame.

2. Smoothing

Smoothing a video means removing the sharpness of the video and providing a blurriness to the video. There are various methods for smoothing such as cv2.Gaussianblur(), cv2.medianBlur(), cv2.bilateralFilter(). For our purpose, we are going to use cv2.Gaussianblur().

3. Edge Detection

Edge detection is a useful technique to detect the edges of surfaces and objects in the video. Edge detection involves the following steps:

- Noise reduction
- Gradient calculation
- Non-maximum suppression
- Double threshold
- Edge tracking by hysteresis

4. **Bitwise Operations**

Bitwise operations are useful to mask different frames of a video together. Bitwise operations are just like we have studied in the classroom such as AND, OR, NOT, XOR.

Neural Network Modeling:Tensorflow(tf.lite), Keras, Pytorch

- **Tensor flow**

There are multiple changes in TensorFlow 2.0 to make TensorFlow users more productive. TensorFlow 2.0 removes redundant APIs, makes APIs more consistent (Unified RNNs, Unified Optimizers), and better integrates with the Python runtime with Eager execution. Many RFCs have explained the changes that have gone into making TensorFlow 2.0. This guide presents a vision for what development in TensorFlow 2.0 should look like. It's assumed you have some familiarity with TensorFlow 1.x.

- **Keras**

Keras is an API designed for human beings, not machines. Keras follows best practices for reducing cognitive load: it offers consistent & simple APIs, it minimizes the number of user actions required for common use cases, and it provides clear & actionable error messages. It also has extensive documentation and developer guides**.**

- **Exascale machine learning**

Built on top of TensorFlow 2.0, Keras is an industry-strength framework that can scale to large clusters of GPUs or an entire TPU pod. It's not only possible; it's easy.

- **Deploy anywhere**

Take advantage of the full deployment capabilities of the TensorFlow platform. You can export Keras models to JavaScript to run directly in the browser, to TF Lite to run on iOS, Android, and embedded devices. It's also easy to serve Keras models as via a web API.

- **Architectures**

Google LeNet, MobileNet, VGG 19 which are pre-trained models on huge datasets with parameters varying through millions together and deep networks.

GANTT CHART

Chapter 5

GANTT CHART

A Gantt chart is a type of bar chart that illustrates a project schedule. This chart lists the tasks to be performed on the vertical axis, and time intervals on the horizontal axis. The width of the horizontal bars in the graph shows the duration of each activity. A Gantt chart is constructed with a horizontal axis representing the total time span of the project, broken down into increments (for example, days, weeks, or months) and a vertical axis representing the tasks that make up the project (for example, if the project is outfitting your computer with new software, the major tasks involved might be: conduct research, choose software, install software). Horizontal bars of varying lengths represent the sequences, timing, and time span for each task. The bar spans may overlap, as, for example, you may conduct research and choose software during the same time span. As the project progresses, secondary bars, arrowheads, or darkened bars may be added to indicate completed tasks, or the portions of tasks that have been completed. A vertical line is used to represent the report date.

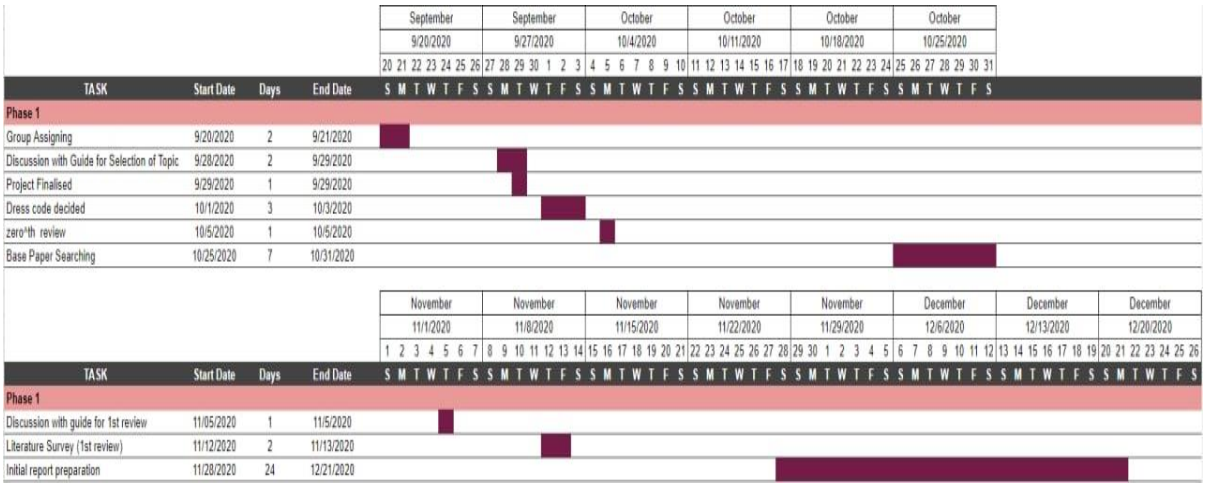

Fig. 5.1: Gantt Chart

REFERENCES

REFERENCES

Reference Papers:

[1] Authors: Yaofeng Xue; Shang Gao; Huali Sun; Wei Qin

"A Chinese Sign Language Recognition System Using Leap Motion"

Published in: 2017 International Conference on Virtual Reality and Visualization (ICVRV)

Link: https://ieeexplore.ieee.org/document/8719155

[2] Authors: Xue Wu; Xiao-ru Song; Song Gao; Chao-bo Chen

"Gesture recognition based on transfer learning"

Published in: 2019 Chinese Automation Congress (CAC)

Link: https://ieeexplore.ieee.org/document/8997255

[3] Authors: Sarfaraz Masood,Harish Chandra Thuwal,Adhyan Srivastava

"American Sign Language Character Recognition Using Convolutional Neural Network"

Published in: In book: Smart Computing and Informatics (pp.403-412)

Link:

https://www.researchgate.net/publication/320703517_American_Sign_Language_Character_Recognition_Using_Convolution_Neural_Network

[4] Authors: Hongyang Gao; Zhengyang Wang; Lei Cai; Shuiwang Ji

"ChannelNets: Compact and Efficient Convolutional Neural Networks via Channel-Wise Convolutions"

Published in: IEEE Transactions on Pattern Analysis and Machine Intelligence (Early Access)

Link: https://ieeexplore.ieee.org/document/9007485

[5] Authors: Qingqing Cao, Niranjan Balasubramanian, Aruna Balasubramanian

"MobiRNN: Efficient Recurrent Neural Network Execution on Mobile GPU"

Published at 1st International Workshop on Embedded and Mobile Deep Learning colocated with MobiSys 2017

Link: https://arxiv.org/abs/1706.00878

[6] Authors: Chirag Patel,Ripal Patel

"Gaussian mixture model based moving object detection from video sequence"

Conference: Proceedings of the ICWET '11 International Conference & Workshop on Emerging Trends in Technology

Link:
https://www.researchgate.net/publication/220902168_Gaussian_mixture_model_based_moving_object_detection_from_video_sequence

[7] Authors: Huy B.D Nguyen; Hung Ngoc Do

"Deep Learning for American Sign Language Fingerspelling Recognition System"

Published in: 2019 26th International Conference on Telecommunications (ICT)

Link: https://ieeexplore.ieee.org/document/8798856

[8] Auhtors: Wang Nan; Zhou Zhigang; Lei Huan; Ma Jingqi; Zhuang Jiajun; Duan Guangxue

"Gesture Recognition Based on Deep Learning in Complex Scenes"

Published in: 2019 Chinese Control And Decision Conference (CCDC)

Link: https://ieeexplore.ieee.org/document/8833349

[9] Authors: Dhruv Rathi

"Optimization of Transfer Learning for Sign Language Recognition Targeting Mobile Platform"

Published as Computer Vision and Pattern Recognition (cs.CV)

Link: https://arxiv.org/abs/1805.06618

[10] Authors: Niels Martínez-Guevara; José-Rafael Rojano-Cáceres; Arturo Curiel

"Detection of Phonetic Units of the Mexican Sign Language"

Published in: 2019 International Conference on Inclusive Technologies and Education (CONTIE)

Link: https://ieeexplore.ieee.org/document/8971399

[11] Authors: Jungheon Lee;Youngsan Koh;Jihoon Yang

"A Deep Learning Based Video Classification System
Using Multimodality Correlation Approach"

Published in: 2017 17th International Conference on Control, Automation and Systems (ICCAS)

Link: https://ieeexplore.ieee.org/document/8204286

[12] Authors: Sanjay Kumar;Manish Kumar

"A Study on the Image Detection Using Convolution Neural Networks and TenserFlow"

Published in: 2018 International Conference on Inventive Research in Computing Applications (ICIRCA)

Link: https://ieeexplore.ieee.org/document/8597204

[13] Authors: Mathieu Virbel1,Thomas Hansen2,Oleksandr Lobunets3

"Kivy – A Framework for Rapid Creation of Innovative User Interfaces"
Published in:
Link:
https://dl.gi.de/bitstream/handle/20.500.12116/8013/Virbel_Hansen_Lobunets_2011.pdf?sequence=2&isAllowed=y
[14] Authors: Ahmed Fawzy Gad
"NumPyCNNAndroid: A Library for Straightforward Implementation of Convolutional Neural Networks for Android Devices"
Published in: 2019 International Conference on Innovative Trends in Computer Engineering (ITCE)
Link: https://ieeexplore.ieee.org/document/8646653
[15] Authors: Soeb Hussain;Rupal Saxena;Xie Han;Jameel Ahmed Khan;Hyunchul Shin
"Hand Gesture Recognition Using Deep Learning"
Published in: 2017 International SoC Design Conference (ISOCC)
Link: https://ieeexplore.ieee.org/document/8368821
[16] Authors: Shivashankara S,Srinath S
"American Sign Language Video Hand Gestures Recognition using Deep Neural Networks"
Published in: International Journal of Engineering and Advanced Technology (IJEAT) ISSN:2249-8958,Volume-8 Issue-5,June 2019
Link: https://www.ijeat.org/wp-content/uploads/papers/v8i5/E7205068519.pdf
[17] Authors: Rahul Chauhan;Kamal Kumar Ghanshala;R.C Joshi
"Convolutional Neural Network (CNN) for Image Detection and Recognition"
Published in: 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)
Link: https://ieeexplore.ieee.org/document/8703316
[18] Authors: Kshitij Bantupalli;Ying Xie
"American Sign Language Recognition using Deep Learning and Computer Vision"
Published in: 2018 IEEE International Conference on Big Data (Big Data)
Link: https://ieeexplore.ieee.org/document/8622141
[19] Authors: Mateen Ahmed;Mujtaba Idrees;Zain ul Abideen;Rafia Mumtaz;Sana Khalique
"Deaf Talk Using 3D Animated Sign Language A Sign Language Interpreter using Microsoft's Kinect v2"
Published in:2016 SAI Computing Conference (SAI)

Link: https://ieeexplore.ieee.org/document/7556002

[20] Authors: Brandon Garcia,Sigberto Alarcon Viesca

"Real-time American Sign Language Recognition with Convolutional Neural Networks"

Published in:

Link: http://cs231n.stanford.edu/reports/2016/pdfs/214_Report.pdf

Reference Websites:

[1] https://www.tensorflow.org/

[2] https://kivy.org/#home

[3] https://www.deeplearning.ai/

[4] https://ieeexplore.ieee.org/Xplore/home.jsp

[5] https://github.com/kivy/buildozer

Reference Text books:

[1] "Machine Learning Yearning", by Andrew Ng

[2] "Machine Learning Techniques ", by Tom M Mitchell

[3] "AI and Machine Learning for Coders: A Programmer's Guide to Artificial Intelligence" by Laurence Moroney

[4] "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems", by Aurelien Geron