JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Study and Implementation of Indian Sign Language Classification

K.Sowmya Reddy Stanley College of engineering and technology for women, Hyderabad, India
S.Sreeja Reddy Stanley College of engineering and technology for women, Hyderabad, India
C.Sanavi Reddy Stanley College of engineering and technology for women, Hyderabad, India

Abstract - Sign language is a visual language that is used by the people who are having speaking disability and listening disability people. Sign language uses body language and manual communication to fluidly convey the thoughts of a person. It is achieved by simultaneously combining hand shapes, orientation and movement of the hands, arms or body, and facial expressions. It can be used by a person who has difficulties in speaking or by a person who can hear but could not speak and by normal people to communicate with hearing disabled people. As far as a deaf person is concerned, having access to a sign language is very important for their social, emotional and linguistic growth. Sign language should be recognized as the first language of deaf people and their education can be proceeded bilingually in the national sign language as well as national written or spoken language. The main aim of this project is to reduce communication barrier between normal people and people with speaking and listening disabilities. In our model we used CNN and RNN for feature extraction classification of words. And also made a study on Dynamic Static word classification. The model is not trained when we gave 56 classes, consisting of 1 video per class. So when we duplicated the same videos 20 times, we got test accuracy of 4%.Then again we increased the data to 40 videos per class, we got test accuracy as 8.7%. So as per our observation, we found that to get a better trained model we need more amount of different data

1. INTRODUCTION

1.1 Introduction

Sign language is manual communication commonly used by people who are deaf. Sign language is not universal; people who are deaf from different countries speak different sign languages. The gestures or symbols in sign language are organized in a linguistic way. Each individual gesture is called a sign. Each sign has three distinct parts: the hand shape, the position of the hands, and the movement of the hands. Deaf people's mother tongue is sign language, which is a visual language. Unlike acoustically transmitted sound patterns, sign language employs body language and manual communication to express a person's thoughts. It can be used by people who have difficulty in listening, people who can hear but cannot speak, and regular people to communicate with people who are deaf or hard of hearing. It is important for a deaf person's psychological, mental, and linguistic advancement to have access to a sign language. Deaf people's first language should be accepted, and their schooling should be conducted bilingually in sign language and written or spoken language. Deaf and hard-of-hearing people use Indian Sign Language to communicate by making various body gestures. There are various groups of deaf people all over the world, and their languages will be different as well. American Sign Language (ASL) is used in the United States; British Sign Language (BSL) is used in the United Kingdom; and Indian Sign Language (ISL) is used in

India for transmitting feelings and communicating. Manual speech and body language (nonmanual communication) are used to express emotions, concepts, and feelings in "Indian Sign Language (ISL)." One-handed, two-handed, and non-manual ISL signals can both be grouped into three groups. Manual signs, such as one-handed and two-handed signs, are made with the signer's hands to communicate information. By altering body posture and facial expressions, non-manual signs are produced. Using sign language, the deaf and the mute can somehow communicate. Only some people understand sign language. Should the need for the deaf or mute to speak publicly arise, people usually employ the help of a translator.

1.2 Problem Statement

Normal people can communicate with each other in their native language and impaired people can also communicate in their native language i.e. Sign Language. If normal people need to communicate with impaired people, they don't understand what others are speaking. These people have to rely on an interpreter or on some sort of visual communication. An interpreter won't be always available and visual communication is mostly difficult to understand. Hence there is a need of the systems which recognizes the different signs and conveys the information to the normal people.

1.3 Objective

The aim of this work is to create a system or model that can recognize alphabets, digits and some static sign language words in Indian Sign Language (ISL) and check the evaluation accuracy.

• And made a study on Dynamic Static word classification.

• The data has been collected from kaggle for alphabets and digits which consists 12000 images for each class, 20 static sign language words which consists of 900 images per class, and study on dynamic sign language words classification.

• For Dynamic sign language classification, we took 56 videos.

2. Literature Survey

We are going to represent literature overview of few papers that we have studied for choosing the topic as follows

J. L Raheja, A. Mishra, A. Choudhary et.al [1] In this paper, A vision-based technique has been used in this paper for hand sign recognition, and a support vector machine (SVM) was used to classify as a sign of ISL. This process consists of the following steps: Capturing video, Skin filtering, Pre-processing, Feature calculation, SVM classification, sign recognition. The captured video was converted to HSV color space for pre-processing and then segmentation was done based on skin pixels. Also, Depth information was used in parallel to get more accurate results. HuMoments and motion trajectory were extracted from the image frames and the classification of gestures was done by Support Vector Machine. In this way, This system targets to facilitates disabled people who are not able to hear and there are not many people who can understand their sign language. The Indian Sign Language Recognition system was implemented and 4 gestures were classified using SVM. Results were 97.5% accurate.

Dongxu Li, Cristian Rodriguez Opazo, Xin Yu, Hongdong Li the Australian National University, Australian Centre for Robotic Vision (ACRV) et.al [2] In this paper they performed two models

Visual appearance-based approach: In This they performed 2D convolution with Recurrent Neural Networks N to capture spatio-temporal features from input video frames. In particular, we use VGG16 [57] pretrained on ImageNet to extract spatial features and then feed the extracted features to a stacked GRU [17]. This baseline is referred to as 2D Conv RNN 2D Human Pose based approach: Pose based approaches mainly utilize RNNs to model the pose sequences for analyzing human motions. RNN to model the temporal sequential information of the pose movements, and the representation output by RNN is used for the sign recognition

Saurabh Kumbhar, Abhishek Landge, Akash Kulkarni, Devesh Solanki, Vidya Kurtadika et.al [3] In this paper they proposed a system for sign language recognition of alphabets and numbers based on CNN. They have used CNN because it increases the accuracy of the system by recognizing hidden patterns and correlation in raw data. This CNN based sign language recognition system, for recognizing the numbers and alphabets of ISL has produced test accuracy of 92.56 % and validation accuracy of 98.34 %Cloud is a router solution for varied retention that tries to give cheap re pairing in the event of an emergency single loss. NC

Canopy is built on top of internet persistence technologies such as renewal codes. For the operationally low storage regenerating code (F-MSR) that keeps the same redundancy, we provide an effective development option. the degree of difficulty.

3. OVERVIEW OF THESYSTEM

3.1 Existing System

• Sign language recognition of alphabets and numbers based on CNN. They have used CNN because it increases the accuracy of the system by recognizing hidden patterns and correlation in raw data.

3.2 Proposed System

In our proposed system our model classifies the alphabets, digits and static words using Convolution Neural Networks (CNN) for feature extraction and model training. We have to pass an image to model then the system recognizes which class it belongs to. And also made a study on dynamic word classification using RNN for model training

3.3 Proposed System Design

Using Deep Learning approach::

1. Data Collection
2. Data Cleaning & Pre-processing text
3. Feature Extraction (CNN)
4. CNN Model Training
5. Testing
6. Classification using CNN
7. Evaluating the accuracy using testing dataset

3.3.1 Data Collection

Collecting data allows you to capture a record of past events so that we can use data analysis to find recurring patterns. It is a very important part of a research work because it enables the researcher to take decisions related to the information available and also to understand how helpful is the information that will assist in carrying forward the research work

3.3.2 Data Cleaning

The main aim of Data Cleaning is to identify and remove errors & duplicate data, in order to create a reliable dataset. This improves the quality of the training data for analytics and enables accurate decision-making.

3.3.3 Data Preprocessing

Real world data is generally Incomplete: Certain attributes or values or both are missing or only aggregate data is available. Noisy: Data contains errors or outliers Inconsistent: Data contains differences in codes or names etc. Tasks in data preprocessing Data preprocessing is an important task. It is a data mining technique that transforms raw data into a more understandable, useful and efficient format

3.3.4 CNN

In deep learning, a convolution neural network (CNN, or ConvNet) is a class of artificial neural network (ANN), most commonly applied to analyze visual imagery.[1] CNNs are also known as Shift Invariant or Space Invariant Artificial Neural Networks (SIANN), based on the sharedweight architecture of the convolution kernels or filters that slide along input features and provide translation-equivariant responses known as feature maps. They have applications in image and video recognition, recommender systems, [5] image classification, image segmentation, medical image analysis, natural language processing, [6] brain–computer interfaces, [7] and financial time series. [8] A convolution neural network consists of an input layer, hidden layers and an output layer. In any feed-forward neural network, any middle layers are called hidden because their inputs and outputs are masked by the activation function and final convolution. In a convolution neural network, the hidden layers include layers that perform convolutions. Typically this includes a layer that performs a dot product of the convolution kernel with the layer's input matrix. Its activation function is commonly ReLU. As the convolution kernel slides along the input matrix for the layer, the convolution operation generates a feature map, which in turn contributes to the input of the next layer. This is followed by other layers such as pooling layers, fully connected layers, and.

3.3.4 RNN:

Recurrent Neural Network (RNN) is a type of Neural Network where the output from previous step is fed as input to the current step. In traditional neural networks, all the inputs and outputs are independent of each other, but in cases like when it is required to predict the next word of a sentence, the previous words are required and hence there is a need to remember the previous words. Thus RNN came into existence, which solved this issue with the help of a Hidden Layer. The main and most important feature of RNN is Hidden state, which remembers some information about a sequence. RNN have a "memory" which remembers all information about what has been calculated. It uses the same parameters for each input as it performs the same task on all the inputs or hidden layers to produce the output. This reduces the complexity of parameters, unlike other neural networks.

4 Architecture

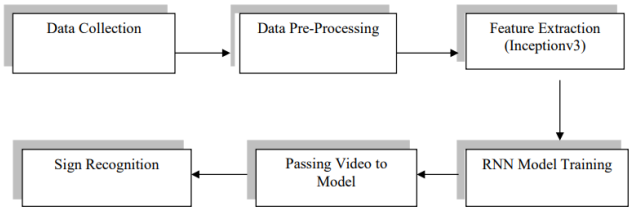

Fig 1: RNN Model

5 RESULTS SCREEN SHOTS

Alphabets and Digits classification The dataset consists of 35 classes, each class consists of 12000 images. Each image is converted into grayscale and resized using Numpy. For model training we used CNN algorithm, it consists of 5 convolution layer, 3 Max Pooling layers, 1 fully connected layer and 5 dropout layers. For activation layers used ReLU and SoftMax. As a result, we got 99% accuracy.

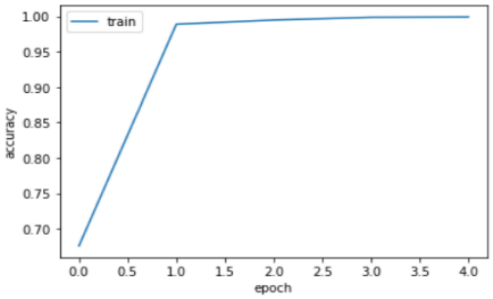

Fig . Accuracy during training process.

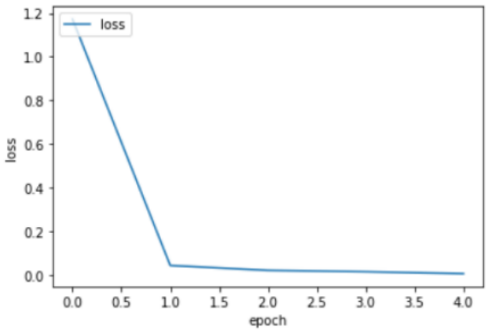

Fig . Loss function during training process.

Static words classification:

The dataset consists of 35 classes, each class consists of 12000 images. Each image is converted into grayscale and resized using Numpy. For model training we used CNN algorithm, it consists of 5 convolution layer, 3 Max Pooling layers, 1 fully connected layer and 5 dropout layers. For activation layers used ReLU and SoftMax. As a result, we got 99% accuracy.

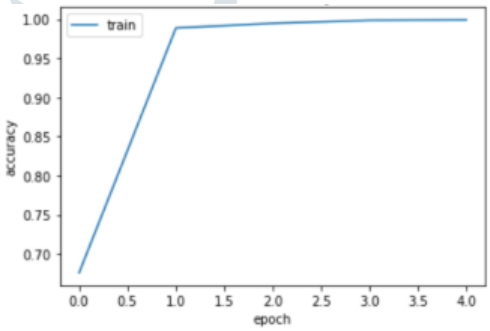

Fig . Accuracy during training process.

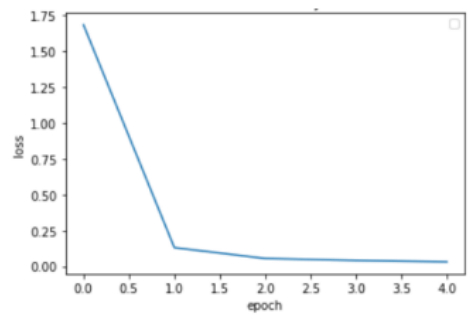

Fig . Loss function during training process.

Dynamic words classification:

At first, we took 1 video per for training a model but as a result the model is not trained. So we duplicated the videos into 20 videos for training as a result we got 4% accuracy so again duplicated the videos into 40 videos and trained it with RNN as a result we got 8.4% accuracy. Later we took 8 classes which consists of 52 videos per class and trained it with RNN as a result we 40.9% accuracy.

6. CONCLUSION

✓ The main aim of this project is to reduce communication barrier between normal people and people with speaking and listening disabilities. In our model we used CNN and RNN for feature extraction classification of words. When we pass an image to our model, it recognizes which class it belongs to. So for alphabets, digit classification and Static words classification we got 99% accuracy. And for dynamic word classification, for 8 classes consisting of 52 videos per class, we got 40% accuracy. So from our study we conclude that, we will get a better trained model only if we have more amounts of different data.

7. References

[1]. Word-level Deep Sign Language Recognition from Video: A New Large-scale Dataset and Methods Comparison

[2]. Systematic Literature Review: American Sign Language Translator

[3]. Indian Sign Language Recognition System

[4]. Sign Language Recognition System, Mayuresh Keni, Shireen Meher, Aniket Marathe.

[5]. Joshi A, Sierra H, Arzuaga E. American sign language translation using edge detection and cross correlation. 2017 IEEE Colombian Conference on Communications and Computing, COLCOM 2017 - Proceedings. 2017.

[6]. CM, Omar Z, Jaward MH. A mobile application of American sign language translation via image processing algorithms. Proceedings - 2016 IEEE Region 10 Symposium, TENSYMP 2016. 2016;: p. 104-109

[7]. Guo D, Zhou W, Wang M, Li H. Sign language recognition based on adaptive HMMS with data augmentation. Proceedings - International Conference on Image Processing, ICIP. 2016; 2016-Augus: p. 2876-288