Sign Language Recognition App

Prof.Vinay Sahu[1], Shristy Nawange[2], Urvashi Dongre[3], Varsha Deshmukh[4], Yogita Bele[5]

[1]Professor, CSE Department, SBITM, Betul, Madhya Pradesh, India
[2,3,4,5]Student, CSE Department, SBITM, Betul, Madhya Pradesh, India

Abstract:
Conversing to a person with hearing disability is always a major challenge. Sign language has indelibly become the ultimate panacea and is a very powerful tool for individuals with hearing and speech disability to communicate their feelings and opinions to the world. Veritably many people understand sign language. It makes the integration process between them and others smooth and less complex. Written communication is time consuming and easy and only well liked when people are stationary. The features extracted are the binary pixels of the images. We make use of Convolutional Neural Network (CNN) for training and to classify the images.

Keywords: Sign Language, Gesture recognition, ASL, Sign language recognition, Convolutional Neural Network (CNN), Machine Learning.

1.INTRODUCTION

Talk to a man in a language he understands, that goes to his head. Talk to him in his own language, that goes to his heart, language is undoubtedly essential to human interaction and has existed since human civilization began. It is a medium humans use to communicate to express themselves and understand notions of the real world. Without it, no books, no cell phones and definitely not any word I am writing would have any meaning. Truly many people understand sign language. Also, polar to well-liked belief, it is not a transnational language. Obviously, this farther complicates communication between the Deaf community and the hail maturity. The volition of written communication is clumsy, because the Deaf community is generally less professed in writing a spoken language.

This type of communication is impersonal and slow in face-to- face exchanges. It is so deeply embedded in our everyday routine that we often take it for granted and don't realise its importance. Sadly, in the fast changing society we live in, people with hearing impairment are usually forgotten and left out. Sign language, although being a medium of communication to deaf people, still have no meaning when conveyed to a non-sign language user. Hence, broadening the communication gap. To prevent this from happening, we are putting forward a sign language recognition system. It will be an ultimate tool for people with hearing disability to communicate their thoughts as well as a very good interpretation for non sign language user to understand what the latter is saying. Many countries have their own standard and interpretation of sign gestures. It'll be an ultimate tool For people with hail disability to communicate their studies as well as a veritably good interpretation for non-sign language stoner to understand what the ultimate is saying. Numerous countries have their own worth and clarification of sign gestures. In this case, we have decided to go with the static recognition of hand gestures because it increases accuracy as compared to when including dynamic hand gestures like for the alphabets J and Z. For case, an ABC in Korean sign language won't mean the same thing as in Indian sign language. While this best part diversity, it also pinpoints the complexity of sign languages. Deep Literacy must be well clued with the gestures so that we can get a decent delicacy. In our proposed system Subscribe Language is used to produce our datasets.

Identification of the sign language can be performed using the grounded system through the dynamic recognition or through glove grounded system. Despite having a delicacy of over 90, wearing gloves are

uncomfortable and cannot be utilized in stormy weather conditions. They are not fluently carried around since their use bears computer as well.

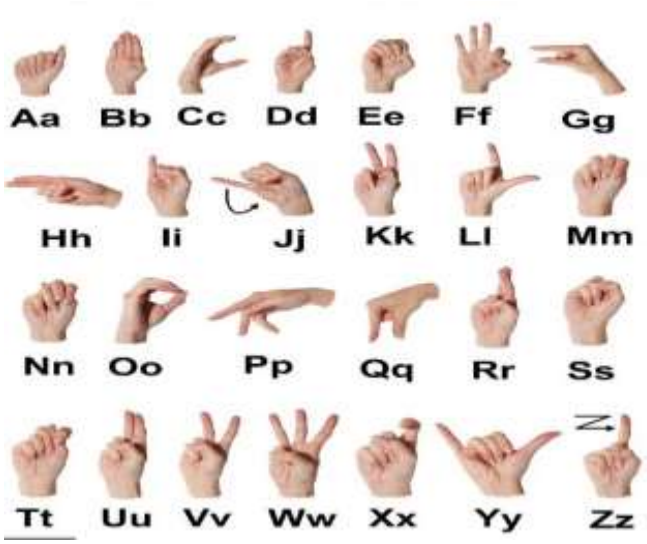


Fig 1. American sign language alphabets

1.1 Objective

The objective of the Sign language Recogntion application is to convert the hand gestures into the visual representation of the word or the sentence over the screen. This will be carried out using Convolutional Neural Network through the use of machine learning algorithms. The main output will be the direct representation of the enacted word onto the screen in real time

1.2 Literature Survey

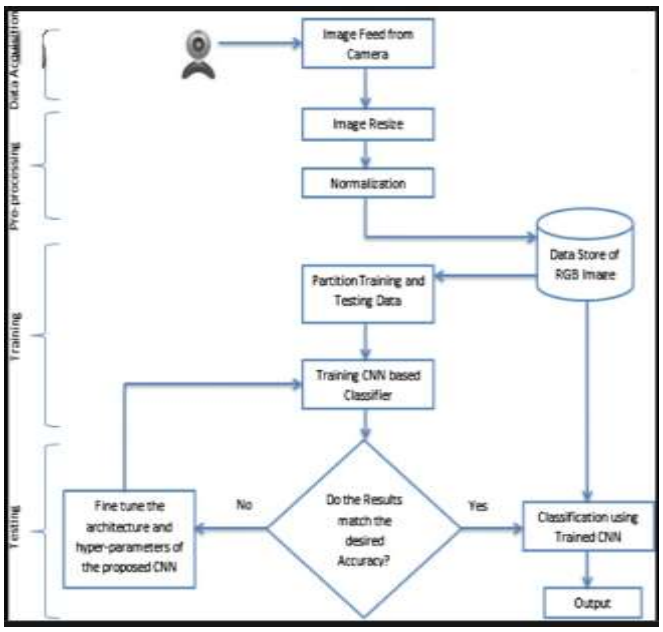
Literature review of our proposed system shows that there have been many explorations done to tackle the sign recognition in videos and images using several methods and algorithms. A system having a dataset of 40 common words and 10,000 sign language images. To locate the hand regions in the video frame, Faster R-CNN with an embedded RPN module is used. It improves performance in terms of accuracy. The discovery delicacy of Faster R-CNN in the paper increases from 89.0% to 91.7% as compared to Fast-RCNN. A 3D CNN is used for point birth and a sign-language recognition frame conforming of long-and short- time memory (LSTM) Rendering and decrypting network are erected for the language image sequences.

On the problem of RGB sign language image or videotape recognition in practical problems, the paper merges the hand locating network, 3D CNN point birth network and LSTM garbling and decrypting to construct the algorithm for birth. This paper has achieved a recognition of 99 in common vocabulary dataset. The image features are extracted and classified with Multi class SVM, DTW and non-linear CNN. A dataset of 23 Indian Sign Language static alphabet signs were used for training and 25 videos for testing. The experimental result obtained were 94.4%for static and 86.4% for dynamic. Make use of 50 specimens of every alphabets and digits in a vision based recognition of Indian Sign Language characters and numerals using B-Spine approximations. The region of interest of the sign gesture is analyzed and the boundary is removed. The boundary obtained is further transformed to a B-spline curve by using the Maximum Curvature Points (MCPs) as the Control points. The B-spline curve undergoes a series of smoothening process so features can be extracted. Support vector machine is used to classify the images and the accuracy is 90%. Convolutional Neural network model having 6 layers for training. It is to be noted that his model is not a 3D CNN and all the kernels are in 2D. He has used Rectified linear Units (RLU) as activation functions. Feature extraction is performed by the CNN while classification uses ANN or fully connected layer.

1.3 Proposed System

The first step of the proposed system is to collect data. Many researchers have used sensors or cameras to capture the hand movements. For our system, we make use of the web camera to shoot the hand gestures. The images undergo a series of processing operations whereby the backgrounds are detected and eliminated using the color extraction algorithm HSV (Hue, Saturation, Value) . Segmentation is then performed to detect the region of the skin tone. Using the morphological operations, a mask is applied on the images and a series of dilation and erosion using elliptical kernel are executed. With open CV, the images obtained are amended to the same size so there is no difference between images of different gestures. Our dataset has 2000 American sign gesture images out of which 1600 images are for training and the rest 400 are for testing purposes. It is in the ratio 80:20. Binary pixels are extracted from each frame, and Convolutional Neural Network is applied for training and classification. The model is then evaluated and the system would then be able to predict the alphabets.

The methodological steps of the Sign Language Recognition system have been described in below-



In this study, the raw images are transformed into grayscale images. The grays levels of input images are normalized by the maximum value of the gray level range. The use of lower solution images provides faster training without too much impact on the recognition rate. The images are resized to 64×64 pixels.

1.4 Process

A. Background Elimination

It is a model which splits the color of an image into 3 separate parts namely: Hue, Saturation and value. HSV is a powerful tool to improve stability of the images by setting apart brightness from the chromaticity. The Hue element is unaffected by any kind of illumination, shadows and shadings and can thus be considered for background removal.

fig (a). Image captured from web-camera.

fig (b). Image after background is set to black using HSV

B. Segmentation

The first image is then transformed to grayscale. As much as this process will result in the loss of color in the region of the skin gesture. Non-black pixels in the transformed image are binaries while the others remain unchanged, therefore black. The hand gesture is segmented firstly by taking out all the joined components in the image and secondly by letting only the part which is immensely conected, in our case is the hand gesture. The frame is resized to a size of 64 by 64 pixel. At the end of the segmentation process, binary images of size 64 by 64 are obtained where the area in white represents the hand gesture, and the black colored area is the rest.

fig (A). Image after binaries.

(B) Image after segmentation and resizing

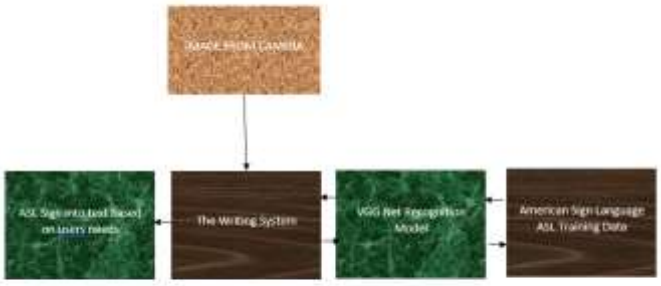
C. Feature Extraction

One of the most crucial part in image processing is to select and extract important features from an image. Images when captured and stored as a dataset usually take up a whole lot of space as they are comprised of a huge amount of data. It also contributes in maintaining the accuracy of the classifier and simplifies its complexity. The features found to be crucial are the binary pixels of the images. Scaling the images to 64 pixels has led us to get sufficient features to effectively classify the American Sign Language gestures. In total, we have 4096 number of features, obtained after multiplying 64 by 64 pixels.

1.5 System Architecture

A **CNN model** is used to extract features from the frames and to predict hand gestures. It is a multilayered feed forward neural network mostly used in image recognition. The architecture of CNN consists of some convolution layers, each comprising of a pooling layer, activation function, and batch normalization which is optional. It also has a set of fully connected layers. As one of the images moves across the network, it gets reduced in size. This happens as a result of max pooling. The last layer gives us the prediction of the class probabilities. The last sub caste gives us the vaticination of the class chances. In our proposed system, we apply

a 2D CNN model with a tensor flow library. The convolution layers scan the images with a filter of size 3 by 3. The dot product between the frame pixel and the weights of the filter are calculated. This particular step extracts important features from the input image to pass on further. The pooling layers are then applied after each convolution layer. One pooling layer decrements the activation map of the previous layer. It merges all the features that were learned in the previous layers' activation maps. This helps to reduce overfitting of the training data and generalizes the features represented by the network. In our case, the input layer of the convolutional neural network has 32 feature maps. of size 3 by 3, and the activation function is a Rectified Linear Unit. The max pool layer has a size of 2×2. The powerhouse is set to 50 percent and the sub caste is smoothed. The last sub caste of the network is a completely connected affair sub caste with ten units, and the activation function is Soft Max. Also, we collect the model by using order cross-entropy as the loss function and Adam as the optimizer.



2. Implementation

Implementation represents the flow of the ordered activities with support for option and iteration.
The Use Case diagram represents the relationship between the user and developer in a sequential order, the order in which interactions occur.

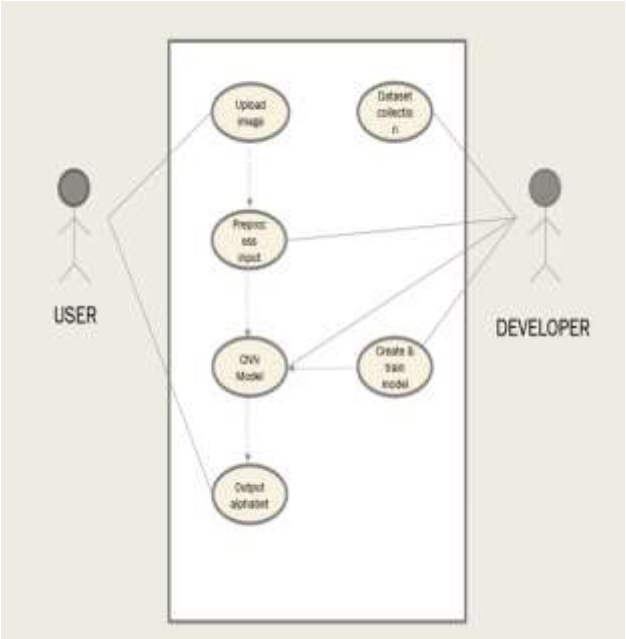


Fig. Use case diagram

2.1 Convolutional Neural

A convolution is a mathematical operation that describes the rule for merging two sets of information. The convolution operation takes input, applies a convolution filter or kernel, and returns a feature map as an output. This operation demonstrates the sliding of the kernel across the input data which produces the convoluted output

data. At each step, the input data values are multiplied by the kernel within its boundaries and a single value in the output feature map is created.

The functional blocks of Convolutional Neural Network:

1. Convolutional Layer
2. Max Pooling layer
3. Fully Connected layer

Convolutional Layer:

It is assumed that the reader knows the concept of Neural networks. When it comes to Machine Learning, Artificial Neural Networks perform really well. Artificial Neural Network are used in various classification tasks like image, audio, words. Different types of Neural Networks are used for different purpose, for example for predicting the sequence of words we use Recurrent Neural Networks more precisely an LSTM, similarly for image classification we use Convolutional Neural Networks.

1. Input Layers:

Input Layer the layer in which we give input to our model. The number of neurons in this layer is equal to the total number of pixel in the case of an image.

2. Hidden Layer:

The input from the Input layer is then feed into the hidden layer. There can be many hidden layers depending upon our model an data size. Ech hidden layer can have different numbers of neurons which are generally greter then the number of features. The output from each layer is computed by matrix multiplication of output of the previous layer with learnable biases followed by activation function which makes the network nonlinear.

3. Output Layer:

The output from the hidden layer is then fed into a logistic function like sigmoid or softmax which converts the output of each class into the probability score of each class.

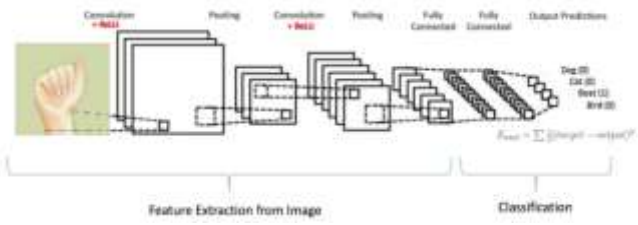
Maximum Pooling or Max Pooling:

Calculate the maximum value for each patch of the point chart. The result of using a pooling sub caste and creating down tried or pooled point charts is a epitomized interpretation of the features detected in the input. They're useful as small changes in the position of the point in the input detected by the convolutional sub caste will affect in a pooled point chart with the point in the same position. This capacity attach by pooling is called the model's invariance to original restatement.

Fully Connected Layer

Completely Connected Sub caste is simply, feed forward neural networks. Completely Connected Layers form the last many layers in the network. The input to the completely connected sub caste is the affair from the final Pooling or Convolutional Layer, which is smoothed and also fed into the completely connected sub caste.

The developed sign language recognition system has been tested on convolutional neural network models. The algorithms with different optimizers are used to train the network for a maximum of 100 epochs with the loss function as categorical cross-entropy. Some of the other parameters which were used to fine-tune the network architecture based upon the preliminary results and after applying some heuristics to increase the accuracy and find an optimal CPU/GPU computing usage are described.



3. CONCLUSIONS

With this complete work, we have studied and learnt various models of machine learning, CNN and image processing which can be used to image classification further. However, most of them require extra computing

power. On the other hand, our research paper requires low computing power and gives a remarkable accuracy of above 90%. In our research, we proposed to normalize and rescale our images to 64 pixels in order to extract features (binary pixels) and make the system more robust. We use CNN to classify the 26 alphabet sign gestures and successfully achieve an accuracy of 98% which is better than other related work stated in this paper.
The major source of challenge in sign language recognition is the capability of sign recognition systems to adequately process a large number of different manual signs while executing with low error rates. For this condition, it has been shown that the proposed system is robust enough to learn 100 different static manual signs with lower error rates, as in contrast to other recognition systems described in other works in which few hand signs are considered for recognition.

REFERENCES

1. https://www.org/research/sign-language-recognitio-system-using-convolutional-neural-network-and-copmuter-vision-IJERTVI20029.PDF
2. https://www.irjet.net/archives/V9/i3/IRJET-V9I3150.PDF
3. https://www.tensorflow.org/lite
4. https://opencv.org/cource-deepplearning-with-tensorflow-and-keras
5. https://www.researchgate.net/publication/35192456_sign_language_recognition_using_convolutional_neural_network
6. https://biblio.ugent.be/publication/5796137/file/5796322.pdf
7. https://towardsdarascience.com/sign-language-to-text-using-deep-learning-7f9c8018c593
8. https://kccemsr.edu.in/public/files/technovision/1/SIGN%20ANGUAGE%20 RECOGNITION%20USING%20 NEURAL%20NETWORKS.pdf
9. https://analyticsindiamag.com/hand-on-guide-to sign-language-classification-using-cnn/
10. https://www.ijert.org/sign-language-recognition-system-using-convolutional-neura-network-and-computer-vision#:~:text=We%20make%20use%20of%20 Convolutional,remarkable%20 accuracy%20of%20 above%2090%25.&text=recognition%2C%20Sign%20 language%20 recognition%202C%Hue%20 Saturation%20Value%20algorithm.