International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 9 Issue 9 ǁ 2021 ǁ PP. 45-49

Sign Language Recognition Using Deep Learning

A.V. Triven Kumar
U17EC128
BIHER
Chennai, Tamilnadu, India

B. Madhu Sashank
U17EC122
BIHER
Chennai, Tamilnadu, India

Ch. Uday Kiran
U17EC147
BIHER
Chennai, Tamilnadu, India

Abstract
Speech impairment bought a communication gap between the 2 class of people and the disabled still struggle at time to express what's inside their them, so to overcome this problem we propose a system where any one Infront of camera can make the gestures and the system will interpret it for others in real time. This uses a deep neural network trained on sign language dataset to predict what the sign actually is and it presents on the screen itself. The prediction time is lesser than in previously developed systems where the images are forward propagated in deep neural network framework.
Keywords – *deep neural network, forward propagated, sign language.*

Date of Submission: 23-08-2021 Date of acceptance: 07-09-2021

I. Introduction

There are roughly 11% of people in USA who have some kind of speech disorder or speech impairment. This includes apraxia of speech, cluttering, stuttering, dysphasia, muteness, speech sound disorder and voice disorder. There are millions of people who cannot afford an interpreter every-time they need to communicate and interpreters are less in number and won't be every-time around. This project intends to fill this gap for unfortunate people to communicate efficiently with others and feel as normal. Depression and low self esteem is sometimes seen among the unfortunate people, so this may help to them overcome not all but few barriers.

Sign Language substantially facilitates communication in the speech impaired community. However, there are only 250,000-500,000 interpreters which significantly limits the number of people that they can easily communicate with. The alternative of sign language - written communication is inconvenient, impersonal and even impractical in numerous situations when an emergency occurs. In order to overcome this hurdle and to enable dynamic communication, we present a Sign language recognition system that uses Convolutional Neural Networks (CNN) in real time to translate a video of a user's ASL signs into text shown on screen.
Our system has mainly 3 steps.

- Taking input video from a device Infront of user.
- Predicting the sign present in each frame of the video.
- Presenting it on the screen for the user.

As this requires a camera to take the video feed and a good hardware to do all the machine learning stuff for prediction it has some limitations.

- Background of user should be consistent.
- Good camera
- Faster hardware.

But the advent of technology, electronic equipment is getting cheaper and better so we believe these limitations can be easily overcome and deployed in a large scale.

II. Related works

Sign language recognition is not a new computer vision problem, people have in past tried to predict the sign language and interpret it using various methods like skin masking Bayesian network, principal component analysis, Douglas -- Peucker algorithm. These require good amount of image preprocessing and highly over trained models. In case of pca, the reduction of dimensions and converting each one the pixel values to pcs removes any information present in them, this cannot be used in any other algorithm effectively.

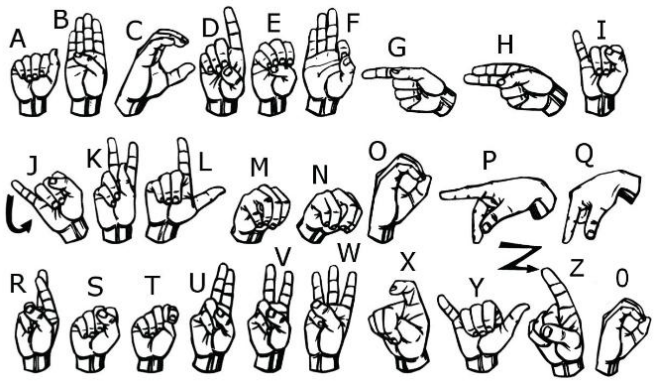
Sift algorithm proved to be the sought after algorithm for everyone in computer vision classification problem but it is still not accurate enough when compared to convolutional neural networks efficiency and feature selection ability.

The problem of alphabet recognition may seem to sound small but the intricacies involved in it cannot be solved using a single algorithm.

III. Dataset

There are numerous sign languages in the world like indo-pak sign language, japnese sign language, American sign language and german sign language. These sign languages are used in their respective regions, all of these various languages have same functionality with few changes in the signs in every language. We chose the American sign language as our dataset as it is readily available in various resources like Kaggle and uci machine learning repository.

ASL – American sign language is widely used in American continent and its dataset is large enough to train a CNN without overtraining on small data. All the alphabets are static in nature (no movement involved) except 'J' and 'Z' in which the gestures are dynamic.



IV. Proposed system

We trained a convolutional neural network of 4 convolutional layers with 32 filters 4 max-pooling layers and 1 drop out layer with 0.3 drop out rate. As the it is a CNN architecture in the activation layer ReLu function has been used which converts the negative values to 0 and only passes positive values. In the output layer categorical cross entropy loss function and Adam optimization functions are used for the model training.

Preprocessing of the image data was done by scaling it to in range 0 to 1 and converting it to a grayscale images, the images are stored in different folders according to their respective labels. Labels column has been encoded using sklearn's lsbel encoder function-converting the string values to numerical values and changing their data type to categorical.

Data augmentation was implemented to ensure the model gets a wide variety of images to be trained on and the predictions hold true in various test environments. The methods used are- horizontal flip, scaling, zoom_range, vertical flip, rotate, brightness and sheer_range.

After training this model for few epochs the model config and its respective weights are saved in local system and loaded in the opencv script using cv-dnn module to read the net. Later video is accessed and some preprocessing is done like converting it to grayscale and then sent to the network frame by frame for prediction. The predicted gesture is then mapped to its respective label and printed on different window.

Fig 2. Flow diagram

We used transfer learning approach in deep learning for the model architecture, in here a predefined architecture is used like vgg, alexnet, or yolo and the final output layer is defined according to the need of usage and pretrained models weights are imported and updated on our custom dataset. The weights on which we trained on was imagenet, ImageNet is quite a beast. It holds 1,281,167 images for training and 50,000 images for validation, organised in 1,000 categories, the size of this dataset is 150 GB. Researchers train on this dataset and get benchmark results and opensource the model architecture and its weighs file, so this can be used by anyone by importing these files, the weights of predefined model are half trained and some features selection is already learnt by it.

A) Tools used

- Python – python is a high level programming language mainly used in data science and web development, its code is user friendly is easy to interpret. We used it in training the model and the computer vision part is done in python using open cv and tensorfow to load the trained model.
- Google colab – google provides a good compute capability cpu along with free gpu of nvidia for deep learning purposes to encourage students, researchers to work in AI field. This platform has been used to train our model in less time.
- Tensorflow – tensorflow is high level deep learning framework developed by google, the version 2.x is built on keras for easy usage and code friendliness. Deploying a deep learning project inn tensorflow is easy due to its javascript version

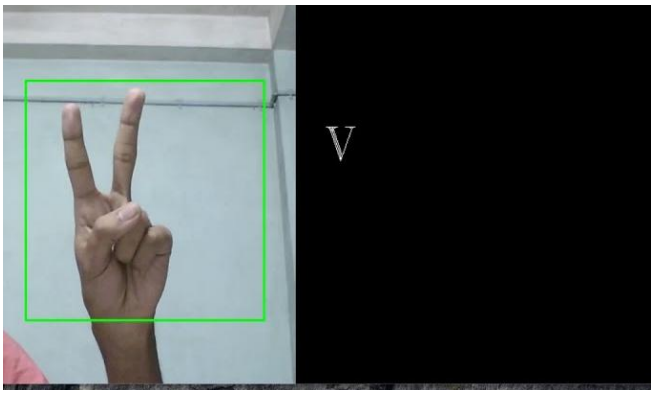
V. Future scope

The implemented system can be deployed in a website using javascript so that it can be used anywhere and it only requires an internet connection to be used. The dataset training took 4 hours in google colab with xenon cpu and tesla t4 gpu, with better hardware this training time can be reduced significantly.

This paper though written for American sign language can be used the same approach to recognize different sign languages too.

VI. Results

Training accuracy achieved for this dataset is 96% and validation accuracy is 92%, as this model has not yet seen the real life scenario where the background is going to change in different places, it still gave better results in different environments. J,Z couldn't be recognized as of this dataset, but an alternative sign for this purpose can be used in transliteration process for easy communication.



References

[1]. Geetha M, Rohit Menon, Suranya Jayan, Raju James, Janardhan G.V.V, "Gesture Recognition for American Sign Language with Polygon Approximation", IEEE International Conference on Technology for Education, pp.241-245, 2011

[2]. Nachamai. M, "Alphabet Recognition of American Sign Language: A Hand Gesture Recognition approach Using Sift Algorithm", International Journal of Artificial Intelligence & Applications, Vol.4, No.1, pp.105-115, 2013.

[3]. Suchin Adhan and Chuchart Pintavirooj, "Alphabetic Hand Sign Interpretation using Geometric Invariance", IEEE International Conference on Biomedical Engineering, 2014.

[4]. Sharmila Konwar, Sagarika Borah, Dr.T Tuithung, "An American Sign Language Detection System using HSV Color Model and Edge Detection", IEEE International Conference on Communication and Signal Processing, pp.743-747, 2014.

[5]. Fahad Ullah, "American Sign Language Recognition System for Hearing Impaired People Using Cartesian Genetic Programming", 5th International Conference on Automation, Robotics and Applications, pp.96-99, 2011.

[6]. Srinath S, Ganesh Krishna Sharma, "Classification approach for Sign Language Recognition", International Conference on Signal, Image Processing, Communication & Automation, 2017.

[7]. B M Chethana Kumara, H S Nagendraswamy and R Lekha Chinmayi, "Spatial Relationship Based Features for Indian Sign Language Recognition", International Journal of Computing, Communications & Instrumentation Engineering, Vol. 3, Issue 2,

ISSN 2349-1469, 2016.
[8]. Asha Thalange, Dr. S. K. Dixit, "COHST and Wavelet Features Based Static ASL Numbers Recognition", 2nd International Conference on Intelligent Computing, Communication & Convergence (Elsevier), pp.455-460, 2016.
[9]. Aran, O., Burger, T., Caplier, A., Akarun, L.: A belief-based sequential fusion approach for fusing manual signs and non-manual signals. PATTERN RECOGN LETTERS 42(5), 812 – 822 (2009)
[10]. Athitsos, V., Sclaroff, S.: Estimating 3D hand pose from a cluttered image. In: Procs. Of CVPR, vol. 2. Madison WI, USA (2003)
[11]. Awad, G., Han, J., Sutherland, A.: A unified system for segmentation and tracking of face and hands in sign language recognition. In: Procs. of ICPR, vol. 1, pp. 239 – 242. Hong Kong, China (2006). DOI 10.1109/ICPR.2006.194
[12]. Ba, S.O., Odobez, J.M.: Visual focus of attention estimation from head pose posterior probability distributions. In: Procs. of IEEEInt. Conf. on Multimedia and Expo, pp. 53–56 (2008).
[13]. Bailly, K., Milgram, M.: Bisar: Boosted input selection algorithm for regression. In: Procs. of Int. Joint Conf. on Neural Networks, pp. 249–255 (2009). DOI 10.1109/IJCNN.2009.
[14]. Bauer, B., Hienz, H., Kraiss, K.: Video-based continuous sign language recognition using statistical methods. In: Procs. of ICPR, vol. 15, pp. 463 – 466. Barcelona, Spain (2000)
[15]. Bauer, B., Nießen, S., Hienz, H.: Towards an automatic sign language translation system. In: Procs. of Int. Wkshp : Physicality and Tangibility in Interaction: Towards New Paradigms for Interaction Beyond the Desktop. Siena, Italy (1999)
[16]. Bowden, R., Windridge, D., Kadir, T., Zisserman, A., Brady, M.: A linguistic feature vector for the visual interpretation of sign language. In: Procs. of ECCV, LNCS, pp. 390 – 401. Springer, Prague, Czech Republic (2004)
[17]. British Deaf Association: Dictionary of British Sign Language/English. Faber and Faber (1992)
[18]. BSL Corpus Project: Bsl corpus project site (2010). URL www.bslcorpusproject.org/
[19]. Buehler P. Everingham, M., Zisserman, A.: Learning sign language by watching TV (using weakly aligned subtitles). In: Procs. of CVPR, pp. 2961 – 2968. Miami, FL, USA (2009)
[20]. Bungeroth, J., Ney, H.: Statistical sign language translation. In: Procs. of LREC : Wkshp :Representation and Processing of Sign Languages, pp. 105 – 108. Lisbon, Portugal (2004)
[21]. Coogan, T., Sutherland, A.: Transformation invariance in hand shape recognition. In: Procs. of ICPR, vol. 3, pp. 485 – 488. Hong Kong, China (2006). DOI 10.1109/ICPR.2006.1134
[22]. Cooper, H., Bowden, R.: Large lexicon detection of sign language. In: Procs. of ICCV : Wkshp : Human-Computer Interaction, pp. 88 – 97. Rio de Janario, Brazil (2007). DOI