International Journal of Engineering, Science and Mathematics
Vol.9 Issue 6, June2020,
ISSN: 2320-0294 Impact Factor: 6.765
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

CONVOLUTIONAL NEURAL NETWORK-BASED SIGN LANGUAGE TRANSLATION SYSTEM

BASEL A. DABWAN

ABSTRACT

Translation System (TS) is a system used by people withe disabilities to contact with other normal persons in the socoty. Since computers are an essential component of our life, the progress of human-computer interaction (HCI) has supported deaf and dumb people. And the important purpose of our suggest system is to progress an smart system which can turn as a translator between normal and deaf or dumb peopel and can be the communication path between people with speaking deficiency and normal people with both effective and efficient ways. The proposed system consists of a Convolutional Neural Network (CNN) based on the deep learning algorithm for effective extraction of handy properties to recognize the American Sign Language (ASL), for classifying the hand sign. This paper constructs to interpret ASL and also gives a complete overview of deep learning-based methodologies for gesture distignwishes. The proposed solution was tested on data samples from ASL data sets and get an overall accuracy of 96.68%. The proposed system was suitable and reliable for Deaf persons. Furthermore, an efficient and low-cost Hand Gesture Recognition (HGR) system for the real-time video stream from a mobile device camera. A separate individual hand gesture is utilized for validation in this article. The proposed system has to be designed with the front of the camera and the output is given in the form of text or audio.

KEYWORDS:

Machine learning;

Hand motion Rcognation;

Sign language;

Image processing;

Convolutional Neural Network (CNN).

Author correspondence:

Basel A. Dabwan,

Doctorate Program,sign language Program Studies

Dr. BabasahebAmbedkarMarathwada University, Aurangabad, India.

1. Introduction

Only many folks acknowledge the that means of the sign. Normally, deaf people area unit denied of normal contact with alternative common persons within the community. HCI is a noteworthy mechanism among people and devices. this can be a noteworthy space of analysis that focuses on the formation and usage of technology and, in specific, integrated interactions among humans and devices. HCI infrastructure has been surprisingly extended and improved by the technical transition[1]. On the grounds of booming usability, rising technologies implement fashionable user interfaces like Non-Touch, Gesture, and Speech recognition. it is a sophisticated and expensive technology to attain. Such recently enforced technologies area unit then incorporated as applying to specific implementations on the grounds of demands and cost-effectiveness. To order to deal with the difficulties, many researchers are attempting to develop these interfaces in terms of performance, accessibility and lustiness [2]. The best style will have several customary options like simplicity, precision, quantifiability, and adaptability. Today, the human gesture is changing into a widespread HCI application, and therefore the utilization of human gestures, that satisfies of these standards, is growing rapidly[3]. The HGR has several uses in varied areas, like video vice, signing recognition and increased reality (SLR). Between these, associate degree SLR is that the best wide utilised technique wherever speech is troublesome. On this angle, this paper suggests a good technique for the extraction of options via CNN. CNN contains of multiple absolutely connected standard layers as a daily multilayer neural network [4]. The design of CNN is intended to manage second pictures adequately [5].Already, CNN has several parameters to coach the pc effectively [6]. Lastly, SoftMax operate is employed to spot the signing gesture. the aim of This paper is to supply the extraction of features and classification technique by CNN. The purpose is to produce associate degree intelligent application with a high degree of accuracy (less computing time usage). wherever a palmy gesture distinguishes time ranges between zero.25 and 0.50 seconds.

2. Literature Review

Many forms of analysis have additionally been conducted on the interpretation of human signs employing a sort of icons. Sign recognition of the letters, though, is harder. most researchers have developed human body-related ways and hand gestures to enhance technology utilization. Kilioz et al. (2015) have enforced associate degree innovative technique for the popularity of dynamic hand gestures on the grounds of time period HCI. They use a six-degree, location huntsman to collect flight knowledge associate degreed depict motions as an orderly series of directional motions in second. solely the motion flight of the hand (except for finger bending and orientation information) is assumed to explain the gestures. The results of the projected technique In respect of gesture identification and recognition potency (73 per cent accuracy) within the flow of motion [7]. Modanwal et al. (2019) eliminated the space between the pc and also the blind by implementing gesture recognition. they're utilizing tactile or bit as simply a replacement to vision. For making gestures, they additionally created a work surface system. Audio input is additionally provided to the user via the earpiece or mike to confirm reliable and economical knowledge input. The prompt dactylology was developed on the premise of a definition about to the Brailler technique accustomed implement the Braille code. The prompt dactylology still needs each hands, however the person needs to place the fingers instead of depressing the buttons. For this experimental work, simply finger-based gestures area unit investigated. A most of thirty one gestures are often made for every hand with the help of 5 fingers. The sign recogni-tion needs to place the fingers instead of depressing the but-tons. For this experimental work, simply finger-based ges-tures area unit investigated. A most of thirty one gestures are often made for every hand with the help of 5 fingers. The sign recognition score of the instructed system was ninety seven.53%. [8]. Josiane Uwineza et al. (2019) instructed a model manage Human-robot interaction. The instructed hand signals recognition utilizing hybrid properties extraction methoed like Hu Moments, color bar chart and Haralick texture, and Extreme Learning Machine (ELM) for classification. The exactness of ninety eight.7% was reached, that is stronger exactness and shows that the ELM approach may be employed in human-robot interactions.[9]. Haria et al. (2017) developed a less hand gesture recognition marker system which will observe each static and dynamic hand gestures. They used a digital camera put in on a laptop com-puter while not the employment of additional cameras or hand markings like gloves. Their system interprets the dis-cover gesture into actions like gap websites and launching applications like VLC Player and Power-Point. various approaches are used for pre-

processing the image, together with algorithms and techniques for noise decrease, edge detection, smoothing, in the course of specific segmentation techniques for boundary extraction, i.e. characteristic the foreground from the background. They used a complete of seven gestures in their gesture recognition system, six of that square measure static, whereas the seventh may be a motion gesture. Contours, and convexity defects were used with a Haar cascade to spot the entity (hand). As applied against some clear background, the gesture detection system was stable and operated with more or less 92.28% accuracy. For situations wherever the background wasn't clear, the accuracy wasn't sturdy at regarding 64.85 percent[10]. Simran et al., (2019), have designed sys-tem needs swish property between individuals and comput-ers within the YouTube app. They introduced 5 gestures to manager numerous functionalities like light-weight, speed, and begin or stop the app. They used the properties of image process, in the course of neural net-works, to categorise the outlined gesture. The exactness for the individual language the signature is 96.03 percent[11] .

3. Proposed Model

The application is organized to get frames from the time pe-riod video stream of camera that numerous techniques of image process are going to be engaged on it. Next, the input frame ought to be remodeled from RGB to Grayscale. to boost the exactitude of the entred gesture, the noise con-tained within the input frame ought to be removed. Besides, the hand phase is discovered from the image, and also the hand sign is extracted from the frame taken. This process, image, i.e. binary picture mode, is then com-pared to the teached model. program is structured to change a client to catch an image from a mobile device camera. Such recorded pictures area unit preserved within the input folder. Then, hand gesture pictures are often wont to support the CNN coaching model. This dataset contains hand gesture alphabets and investigating digits pictures. There area unit two thousand photos per category, i.e. for every letter and digit. After that, the pictures obtained area unit wont to train the CNN model. In this, the imag-es, that area unit remodeled antecedently into binary mode, area unit fed to the CNN model for process. That Eighty present of the info was provided for coaching and common fraction for the info given for testing. within the coaching output, the model generates a file of kind h5 that stores an outline of the coaching method. By the model, this file are going to be used for prediction of alphabets and digits. Eventually, the user input image is provided to the CNN model for prediction, that compares the input pictures and also the pictures recorded within the CNN model. looking on a comparison, the CNN model generates output in text or

audio format. the essential principle of our model is projected to classify the ASL alphabet, supported the human hand gesture. The operative methodology for the prompt model is seen in Figure1.

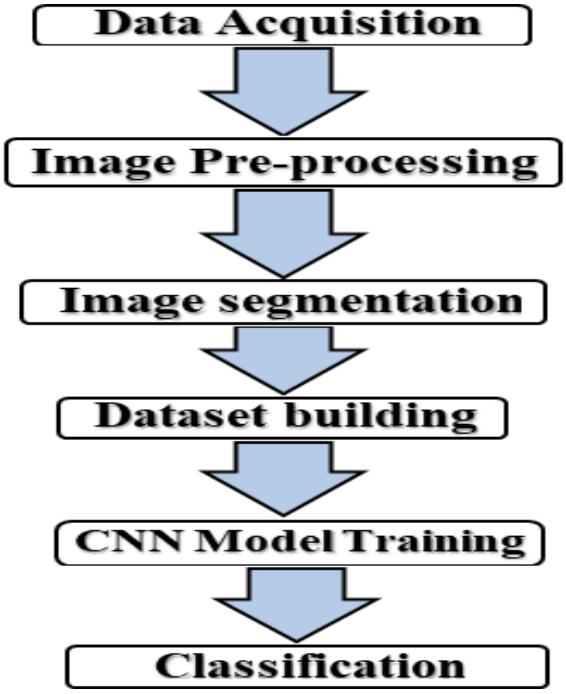

Figure 1: Operational process of the suggested mode

3.1 Data Acquisition

First of all, take a hand gesture video in time period from a hand-held laptop, i.e. Tablet, and acquire the video frames for more calculations.

3.2 Pre-processing stage

At now. The region of interest (ROI) is separated from a frame that's known on a border basis and is outlined by one frame marker. the worth of the grayscale intensity is specified and wont to divide the region into foreground and background areas reckoning on the pel intensity. Intensity values of zero (or background) or one (for foreground) are often appointed to the pixels within the frame. once it categorizes the region as a hand or background, disregards the unused remainder of the video frame, and resizes the frame to a selected resolution. The region of interest as shown in Figure two. Afterwards.

Figure 2: ROI

3.3 Segmentation and feature extraction stage

We have one in all 2 modes of pre-processing the acquired pictures. Binary Mode is employed for the plain background to convert the picture to grayscale, wherever Outso approach is employed to convert the gray picture to the binary pictures wherever the worth of either zero (for background) or one (for foreground). Whereas background subtraction is employed for the compound background to induce a set of a picture that it calculates the arithmetic of the front mask between the buffer store and therefore the background picture, that is, the stationary portion of the sequence or, additional typically, something that may be outlined because the background in keeping with the propperties of the scene being studied[12].In every of those, more noise removal techniques like Gaussian blur, Erosion ar applied. Morphological filtering is vital to use morphological filtering to segmental pictures to make a cleaner, additional closed and additional contoured gesture. this is often accomplished by a series of abrasion operations over the turning of the invariant segmental gesture picture.

3.4 Customized Dataset

HGR screenshots for twenty six letter signs ar obtained for American Sign Language, 10 enumeration digits, and 3 distinctive characters for 3 distinct individuals. There are a 2000 pictures for every sign and human. There ar 3x2000 x pictures (26 alphabets + 10 numbers + 3 characters). Afterwards, frames ar processed as pictures in folders. The folder name is getting used to mark the photographs consistent with the class of gestures within the frame (i.e. mark A for alphabet A, and then on for sure classes of gestures). everey illustration of the language will be used, the interpretation projected is for the yankee language Alphabet, see Figure three.

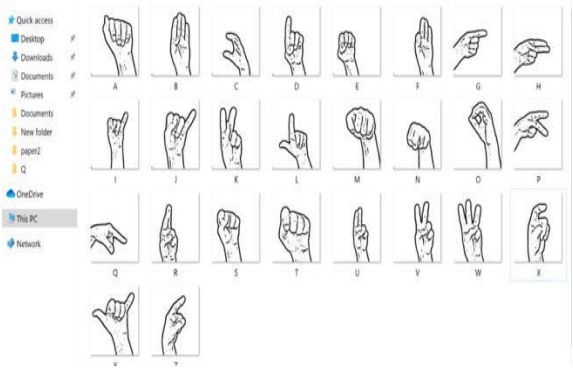

Figure 3: ASL dataset Images

3.5 Feature selection &Training stage

At this time, the extracted properties chosen for classification. For this urged model of feature extraction, the vector component springs from the frame of a video sequence using by the CNN. several of the extracted properties of the image square measure unbroken once extraction within the file. Powerful algorithms of machine learning square measure accustomed extract properties. one amongst the strongest deep learning ways is CNN. A broad of various pictures varies that CNN is employed. CNN can collect potential features for the classification model across a large form of photos. The planned network is four hidden layers. The input dimension is specific for the primary layer. during this paper, the resolution of the illustration of the information is 300x300x1. See Figure four.

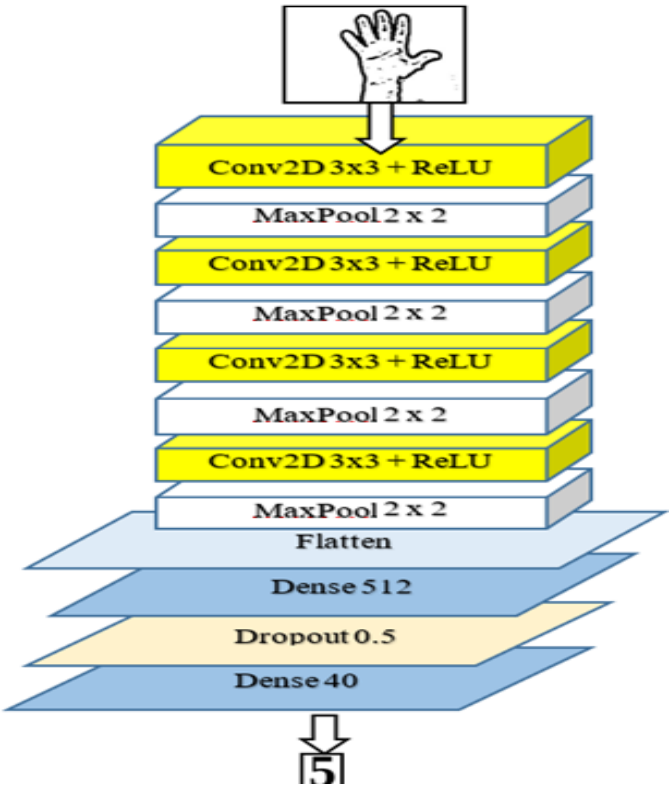

Figure 4: CNN architecture

CNN generates associate degree activation type, that is employed as associate degree input picture for each layer. the method continues layer when layer. Indeed, there area unit simply many layers applicable for the extraction of image features throughout CNN. within this prompt model, hidden layers area unit assumed because the extraction of properties. With these layers, the easy image properties of the layers area unit obtained at the network started. Deeper layers of the network method these essential features and merge the preliminary feature to construct image features of the upper level. These area unit all properties at the next level area unit well-designed for categorization activities. Since deeper layers of the network incorporate of these features of rudimentary into a higher picture exemplification.

3.6 Classification stage

Finally, the SoftMax operate is employed to quantify each alphabetical sign throughout the ultimate a part of this urged model. that's a generalization of supply regression in so far because it will be enforced to continuous information (rather than binary classification) which might embody several decision boundaries. It deals with multinomial labeling

mechanisms. Softmax is that the operate that we tend to continuously take into account the output layer of the classifier. The softmax activation operate provides the distribution of the chance between reciprocally exclusive output categories. SoftMax is often utilised as a tool of learning for the outline of derived regression and features. SoftMax may be a classifier supported a supervised learning technique for categorizing information in many categories. The central operational technique of SoftMax is to assign the sample data input into several distinct classes. In SoftMax, one class is differentiated from the other and a final decision is made for this procedure by selected the maximum output SoftMax value.

4. Result and Evaluation

Python with the Keras and TensorFlow backend libraries was wont to apply the steered deep learning algorithm. The steered model is evaluated through associate interconnected data assortment composed of forty symbols of 3 distinct people. every sign consists of 2000 pictures of every individual. So, in all, there are 3*2000*40 pictures. The dataset is split into 2 teams. the primary branched assortment includes eighty p.c for training pictures and therefore the other includes the rest of the twenty p.c for testing pictures.

For the extraction of a feature that CNN is employed. when victimization CNN for the feature extraction of the photographs, then, we consider the number of teaching features and the number of examining features.. Those are all properties are helpful, that is contributing to distinguish the classes in each individual sign. See figure 5 that show model accuracy about 98 % in an average.

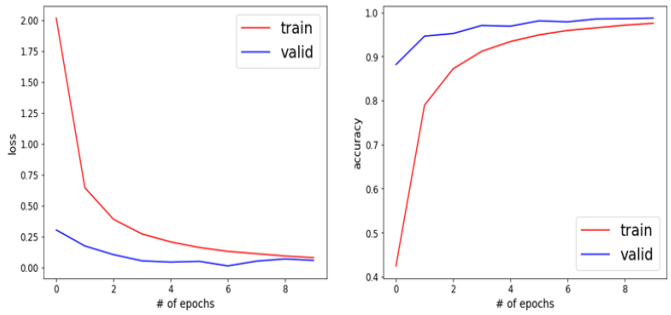

Figure 5: Graph of loss & accuracy against number of epochs for proposed model.

With these more detailed features of training and testing, each of the signs individually presented by the SoftMax has been identified. The quality of the classification has been accepted, which is 96.68%.

Table 1: Comparing the state of the art with the suggested approach

Author	Method	Accuracy	Static/dynamic gestures recognition	No. of recognized gestures	Processing time
Kilioz et al. (2015) [7]	A six-degree position sensor for gathering trajectory data and representing motions as an ordered sequence of spatial motions in 2D.	73%	both	10	1.50 sec
Haria et al.(2017) [10]	Background subtraction to find contours, and convexity defects were used with a Haar cascade to identify hand	92.28%	both	7	Not reported
Modanwal et al. (2019) [8]	dactylology based on a concept similar to the Brailler concept	97.53 %	static	31 gestures of each hand	Not reported
Josiane Uwineza et al. (2019) [9]	Hu Times, the structure of Haralick, and the color histogram for the extraction of features. Extreme Learning Machine for Categorization (ELM)	98.7%	static	9	109.7 s
(Simran et al. (2019)[11]	Image processing features(Image Preprocessing, Segmentation, Gaussian blur, Morphological filtering), followed by CNN for classification	Val acc:98.98%, Test Acc:96.03 %	static	5	Not reported
Our Proposed System	threshold process, background subtraction to detect object and CNN for feature extraction and classification	Val acc:98.11%, Test Acc:96.68	static	40	0.50 sec.

5. Conclusion

Hand sign recognition may be a massive downside in real-life implementations regarding the exactness and responsibility related with it. This paper introduces the hand sign identification in ASL while not touching, the input gestures square measure recorded using a mobile device camera. A still -hand shot taken from a period of time video stream frame and victimization CNN to go looking for additional perceptive features. Eventually detection the sign of the alphabet by SoftMax. Regarding the validity of the model that's planned, our designed dataset is employed in compliance with the ASL conventions. Classification accuracy achieved 96.68%, that is notable with the implementation of ASL signing Recognition with Disabled Persons as the production of HCI. The gesture has got to be presented ahead of the camera and therefore the output is given within the kind of text or audio.

References

[1] *J. F. and H. H. Jonathan Lazar, Research Methods in Human Computer Interaction. Morgan Kaufmann, 2017.*

[2] *A. Mantri and M. Ingle, A Comparative Study of Various Techniques Used in Current HGRSs. Springer Singapore, 2018.*

[3] *R. A. Bhuiyan, A. K. Tushar, A. Ashiquzzaman, J. Shin, and R. Islam, "Reduction of Gesture Feature Dimension for Improving the Hand Gesture Recognition Performance of Numerical Sign Language," pp. 22–24, 2017.*

[4] *B. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," 2012.*

[5] *A. Z. Karen Simonyan, "VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION," pp. 1–14, 2015.*

[6] *J. Donahue et al., "DeCAF : A Deep Convolutional Activation Feature for Generic Visual Recognition," vol. 32, 2014.*

[7] *U. G. Nurettin Çag˘rı Kılıboz, "A hand gesture recognition technique for human–computer interaction," vol. 28, pp. 97–104, 2015, doi: 10.1016/j.jvcir.2015.01.015.*

[8] *G. Modanwal and K. Sarawadekar, "Utilizing gestures to enable visually impaired for computer interaction," CSI Trans. ICT, no. June, 2019, doi: 10.1007/s40012-019-00251-w. and Y. J. , Hongbin Ma, Baokui Li, Static Hand Gesture Recognition for Human Robot Interaction. Springer International Publishing, 2019.*

[9] *A. Haria, A. Subramanian, N. Asokkumar, S. Poddar, and J. S. Nayak, "Hand Gesture Recognition for Human Computer Interaction," Procedia Comput. Sci., vol. 115, pp. 367–374, 2017, doi: 10.1016/j.procs.2017.09.092.*

[10] *P. P. M. C. Simran Shah, Ami Kotia, Kausha Nisar, Aneri Udeshi, "A Vision Based Hand Gesture Recognition System using Convolutional Neural Networks," Int. Res. J. Eng. Technol., vol. 463, no. 6, pp. 2570–2575, 2019, doi: 10.1007/978-981-10-6571-2_132.*

[11] *N. Umadevi and I. R. Divyasree, "Development of an Efficient Hand Gesture Recognition system for human computer interaction," no. September, 2018, doi: 10.18535/Ijecs/v4i12.5*