Recognition of American Sign Language using Deep Learning

Dr. Kiran Babu T S, Dr. Manoj Challa
Department of CSE, CMR Institute of Technology, Bengaluru.

Abstract— Speech impairment is a condition that affects a Person's ability to communicate via voice and hearing. People who are afflicted employ alternative communication methods such assign language. In Spite of the fact that American Sign Language has become more frequently applied in recent years, non-sign language speakers and signers continue to have problems communicating. Thanks to the advancements 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 intention of this paper is to create a vision-based program that converts sign language to text, letting signers and non-signers communicate more easily.

Keywords—deep learning, ASL recognition system, Classification, real-time convolutional neural networks (CNNs)

I. INTRODUCTION

In the deaf community, American Sign Language (ASL) is a significant mode of communication. However, there are only 250,000-500,000 ASL speakers, limiting the number of persons with whom they can interact readily. When an emergency comes, textual communication is inconvenient, impersonal, and even impractical. We describe an ASL recognition system that converts a video of a user's ASL signs into text remove this barrier and enable dynamic communication.

1. Taking video of the user using sign language as an intake (input).
2. Assigning a letter to each frame of the footage
3. Using categorization scores to reassemble and present the most likely term (output).

Neural networks, on the other hand, have previously been used to recognize ASL letters (Appendix A) with accuracies of over 90% [2-11], Only one of them allows real-time classifications, and several of them require a 3-D capturing device such as motion-tracking gloves or a Microsoft Kinect. The scalability and viability of these solutions are hampered by the additional requirements. A pipeline in our system accepts video shows a user signing a word using a web application as input. We then extract individual frames from the video using a CNN and generate letter probabilities for each one (letters a through y, excluding j and z since they require movement).We employ a variety of heuristics to organize based on the frames their content...

II. RELATED WORK

There have been a lot of approaches to recognizing motions using finger writing, but

they have all been limited in the amount of detection rate and time. [2] proposes a categorization strategy for sign language recognitions. This system detects 24 AS alphabet motions and has a success rate of 86.67 percent. A real-time ASL identification system featuring 26 English alphabets with complex backdrops and mixed lighting conditions was demonstrated with an 88.26 percent success rate using Edge Oriented Histogram [11] and a 10 Megapixel we camera with a limiting distance of 1 meter. Mahesh Fernando et al. [12] introduced a recognition system. From a pool of 50, ASL signed gestures, five were chosen by each of the ten signers (A, B, C, D, and V Signed gestures). A total of eight signs were stored and removed as masters containing the recognized five and three more signed motions (A, B, C, D, L, P, V, and Y). In a typical backdrop, 12 signed gestures were unable to be recognized, resulting in a recognition rate of 76 percent utilizing Hu moment categorization. In [13], a Self-Organizing Map was used to present an ASL Recognition System. In a real-time situation with a plain background and a set from the internet, a total of seven different ASL gestures (B, C, H, I, L, O, Y) from ten different sets were examined, with a recognition rate of 92 percent. In 2011, a Cartesian Genetic Programming-based ASL Recognition system was created to recognize the movements of the 26 ASL English alphabet gestures [14In this approach, there are 26 gestures for training and a fresh set of 26 gestures for recognition. More than 90% of the recognition results were true. In [15], the Statistical Measures Technique, Orientation Histogram Technique, COHST (Combined Orientation Histogram and Statistical Technique), and Wavelet Features Technique were used to recognize static ASL signs of numbers 0 to 9 in a plain background, and the recognition rates were 74.69 percent, 82.92 percent, 87.94 percent, and 98.17 percent, respectively..

The ASL Numbers were identified using an Open-Finger Distance Feature Measurement and Neural Network Classification Technique in [16], with a recognition rate of 92.09 percentile. In 2014 [17], a method for identifying ASL was created by combining the HSV color model with an edge detection methodology and morphological techniques to identify human skin color. A total of 100 gestures were investigated, with 65 percent accurately identifying them. Sruthi Upendran et al. built an ASL interpreter [18] that recognized 24 static ASL alphabets in textual form and then converted them to speech with a recognition rate of 77.29 percent using principal component analysis (PCA) and the K-Nearest Neighbor (KNN) approach. Using grey scale thresholding and edge detection techniques, [19] made a human interaction system that could recognize the ASL gesture 'P' against a blank background. They have simply made one gesture to be taken into account. The standardized ASL, which consists of 26 American manual alphabets from A to Z, has been recognized using the MAdaline Neural Network classification technique [20]. Using the Douglas-Peucker algorithm method and polygon approximation, an unique technique for recognizing the 26 static ASL gestures (A-Z) has been developed. This approach accurately recognizes open and closed finger movements with 79.92 percent accuracy [21]. The SIFT technique is used to develop an ASL space, size, illumination, and rotation invariant letter recognition approach [22]. This solution can work with both standard and customized ASL databases. In addition, a quantitative effort to recognize real-time gestures is underway. A dynamic simple and advanced backdrop hand gesture recognition (HGR) integrated system is created using

Gaussian and canny filters with a flood fill approach [23].The letters A through L are assessed for recognition, and in simple and complex backgrounds, they offer 84 percent and 58 percent accuracy, respectively. Background subtraction and finger segmentation techniques are used in an HGR method presented by Zhi-hua Chen et al. [24]. To predict the gesture labels, the rule classifier is employed. The findings are enhanced after 1300 hand movements are used to measure the performance. Using the B spline curvature idea and geometric invariance method, [25] uses a hand gesture interpretation methodology to recognize 24 ASL alphabets movements captured by web camera, with successful results. [26] An ASL finger spelling recognition system is created using phonological feature-based tandem models and a Gaussian mixture observation distributions- based Hidden Markov Model (HMM) baseline In a studio setting with two signers, experiments on finger spelling word recognition are carried.

III. PROBLEM STATEMENT AND METHODS

Due to a multitude of factors, including environmental problems (e.g. lighting sensitivity, background, and camera position), this topic poses a considerable difficulty in terms of computer vision. Occlusion is the absence of something (For example, part or all fingers, or an entire hand, may be out of sight), Co-articulation
(where a sign is impacted by the preceding or subsequent sign), Sign border detection can include abbreviations and acronyms.

SYSTEM ARCHITECTURE

Using CNNs algorithm with real colored photos, a real-time ASL finger spelling recognition was constructed in this study. There were 26 alphabets in total, namely J and Z, as well as two delete and empty classes. The data gathering phase is the first of three phases that make up this system. Since the Hand-Gesture Recognition algorithms studied in this study required a large dataset for training, additional datasets with a bigger range of features, such as various lightings, skin tones, backgrounds, and scenarios, were produced. In the second stage, which was a multi-class recognition with CNN, the writing system symbolized communication between the machine and the user. This method makes it easier for people from both hearing and deaf cultures to communicate. It's a computer-based input system that makes use of a camera. The design of the building.

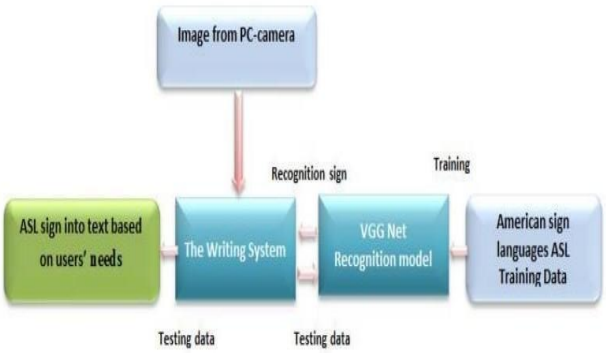

Figure 1. System architecture

Figure 1 demonstrates the architecture of the proposed system..

MULTI-CLASS-RECOGNITION WITH CNNS

This system differed from the majority of earlier works, which used feature-based methodologies. Convolutional neural networks (CNNs) and a deep learning data structure were used to develop a multi-class recognition system. Each ASL sign was divided into its own category. The output of the classifier would be divided into one of 28 groups, ranging from 0 to 27. VGG Net [17], for high-scale image recognition, an incredibly deep convolutional network design was used as the CNN architecture. The suggested re-training approach began by randomly initializing the network weights, and then adjusted the weights to accomplish the tasks with less errors. The network's weights were preserved and loaded as the beginning weights for further tests, a process called as fine-tuning. VGG blocks from the original TFlearn (TFlearn Development Team Github, 2017) available were used in this work.

RECOGNITION SYSTEM FOR MULTI-CLASSES TRAINING

In American Sign Language, the visuals were used to create 28 classes of static fingerspelling (ASL).To feed the VGG Net, all of the photos were resized to 224 by 224 pixels and then normalized. The path of the photographs, along with each one's label, was saved to a text file. NumPy arrays were created from the photos (No. of Images, 32, 32, 3) and fed to the system using TFlearn data. The model was trained with a total of 61.614 training datasets, with at least 2200 for each class. About 0.30 of the training datasets were used for the validation datasets, resulting in a total of 43,120 training datasets and 18,480 validation datasets. The photos were transformed into NumPy arrays (number of images, 32, 32, 3) and supplied into the system via TFlearn data. The model was trained with a total of 61.614 training datasets, with at least 2200 for each class.

IV RESULTS

A CNN algorithm was used to build an ASL identification using real colored images from a PC camera. Text statements are generated from deaf signs in this study assisting in the creation of a writing system that may be

utilized as an input method for any computer equipped with a camera. Using a deep learning technique, this system produced excellent results. The experiment yielded a large number of outcomes. VGG Net was used to create a multi-class recognition system. Each ASL sign was denoted by the letters a unique property using CNNs as the recognition system. The classifier's outcome would be one of 28 categories ranging from 0 to 27. The system recognized 10 ASL letters in the early stages: A, B, C, D, E, F, G, H, I, J. To create less than 1995 pictures, the system was trained with only 10 labels and approximately 20300 training data for each class.

Our notable factors have been the CNN models that have been implemented. Our top model has a validation accuracy of 99.70 percent (0.30 percent error rate). Furthermore, ReLUs have a 23.8 percent improvement over tanh units, indicating that they are quite effective. The accuracy on the primary test set is 99.68%, with a 0.32 percent false positive rate due to noise movements. Because the validation set does not include the users and backdrops that are present in the training set, the test result is

higher than the validation result.

.

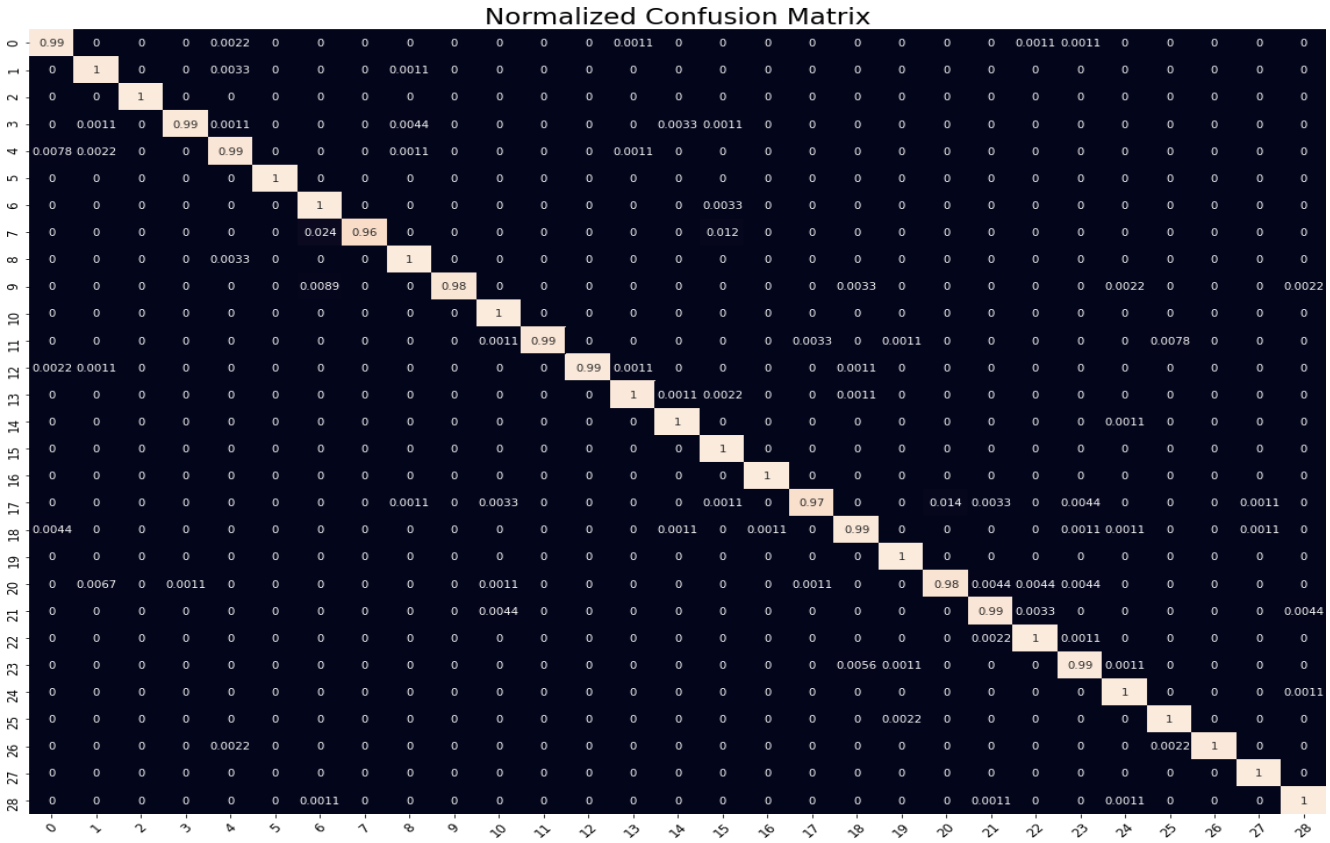

Figure 2. Result-Normalized Confusion Matrix

Experimental Evaluation The CNN model is trained with the MNIST ASL dataset. The data set of 27455 training samples of 784 features is used to train the model. The model was trained to minimize loss by usage of cross entropy ADAM [5]. The various model is trained for 10 epochs on a batch size of 128. The model was trained with a learning rate of 0.001 with 0 decay. The validation dataset consists of 7172 samples, reportedly the validation accuracy of the model is greater than 93 %. The following are the training accuracy & loss plots..

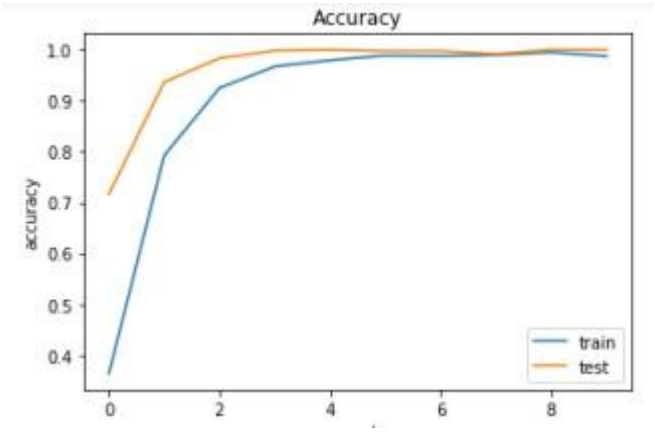

Figure 3: Metric Plot of CNN Model

V.CONCLUSION

The goal of this study is to demonstrate how convolutional neural networks can be used to reliably identify different signs in a sign language without include individuals or their environment in the training set. This general capability of CNNs on spatiotemporal data could aid studies in automatic sign language recognition. When we consider all of the conceivable combinations of motions that a system like this must interpret and translate, sign language recognition is a difficult challenge. Having said that, the best way to solve this challenge is to break it down into smaller chunks, with the system provided here serving as a possible answer to one of them. Despite its poor performance, the system demonstrated

that a first-person sign language translation system could be constructed utilizing only cameras and convolutional neural networks. The model was discovered to have a habit of mixing

up various signs, such as U and W. However, after some thought, flawless performance may not be required because the The employment of a word predictor or an orthography corrector may improve translation accuracy. The next step is to evaluate the response and come up with methods to improve it. More high quality data, additional convolutional neural network topologies, and a reconfigured vision system could all help..

References:

1. Ira Cohen, Nicu Sebe, Ashutosh Garg, Lawrence S Chen and Thomas S. Huang, "Facial expression recognition from videosequences: temporal and static modeling", *Computer Vision and Image Undertaking*, pp. 91, February2003.
2. Bernard Boulay, Francois Bremond and Monique Thonat, "Human Posture Recognition in Video Sequence", *IEEE International Workshop on VSPETS Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, 2003.
3. Anup Nandy, Jay Shankar Prasad, Soumik Mondal, Pavan Chakraborty and G. C. Nandi, "Recognition of Isolated Indian Sign Language Gesture in Real Time" in Communications in Computer and Information Science book series, CCIS, vol. 70.
4. Neha Baranwal and G. C. Nandi, "Continuous dynamic Indian Sign Language gesture recognition with invariant backgrounds by Kumud Tripathi", *2015 Conference on Advances in Computing Communications and Informatics(ICACCI)*.
5. Carol Neidle, Ashwin Thangali and Stan Sclaroff, "Challenges in Development of the American Sign Language Lexicon Video Dataset (ASLLVD) Corpus", *5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon LREC 2012*, 2012.
6. Szegedy et al., "Going Deeper with Convolutions", *CVPR 2015*.
7. Sepp Hochreiter et al., "Long Short-Term Memory", *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.
8. Diederik P. Kingma and Jimmy Ba, "Adam: A Method for Stochastic Optimization", *Published as a conference paper at the 3rd International Conference for Learning Representations*, 2015.
9. Geoffrey Hindon et al., "Dynamic Routing Between Capsules", Nov 2017.

10. Shuai Li, Wanqing Li, Chris Cook, Ce Zhu and Yanbo Gao, "Independently Recurrent Neural Network (IndRNN): Building A Longer and Deeper RNN", CVPR 2018.
11. Maahin Rathinagiriswaran, Swapneel Managaokar, K R Yashaskara Jois, Kartik Vijaykumar Suvarna, Niranjana Krupa, "Inflated 3D Architecture for South Indian Sign Language Recognition", 2021 IEEE International Conference on Mobile Networks and Wireless Communications (ICMNWC), pp. 1-6, 2021.
12. Wasupon Phothiwetchakun, Thanawin Rakthanmanon, "Thai Fingerspelling Recognition Using Hand Landmark Clustering", 2021 25th International Computer Science and Engineering Conference(ICSEC), pp. 256-261, 2021.
13. Nemil Panchamia, Jay Mehta, Priyesh Ghosh, Jalpa Mehta, "ASL Tutor Using Deep Learning", 2021 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON), pp.1-6, 2021.
14. Sirisha M, Sanjana B, Vishal Goutham N, Surekha Borra,"Evaluation of Machine Learning Models for Real-Time Sign Recognition", 2021 IEEE Mysore Sub Section International Conference (MysuruCon), pp. 238-243, 2021.
15. Sadaf Ikram, Namrata Dhanda, "American Sign Language Recognition using Convolutional Neural Network", 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), pp. 1-12, 2021.
16. Mohammad Daim Khan, B Shivalal Patro, Rajiv Ranjan, Manas Chandan Behera, Raushan Kumar, Utsav Raj,"Real-Time American Sign Language Realization Using Transfer Learning With VGG Architecture", 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), pp. 1-5, 2021.
17. Ashwin Kannoth, Cungang Yang, Manuel Angel Guanipa Larice, "Hand Gesture Recognition Using CNN & Publication of World's Largest ASL Database", 2021 IEEE Symposium on Computers and Communications (ISCC), pp. 1-6, 2021.