International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 05 | May - 2022 Impact Factor: 7.185 ISSN: 2582-3930

IDENTIFICATION OF AMERICAN SIGN LANGUAGE TO SPEECH CONVERSION USING TRANSFER LEARNING (SHUFFELNET)

Vijay J[1], Vijayakumar J[2], Mervin Paulraj M[3]

[1]Student, Department of Electronics and Instrumentation, Bharathiar University, Coimbatore
2Associate Professor & Head, Department of Electronics and Instrumentation, Bharathiar University, Coimbatore
[3]Research Scholar, Department of Electronics and Instrumentation, Bharathiar University, Coimbatore

---

Abstract - Sign language is used to facilitate communication between normal people and deaf and hard of hearing people. Many people are unaware of sign language, which creates a barrier to knowledge sharing and comprehension of thoughts and feelings. The convolution neural network is used in this study to train sign languages using the ShuffleNet transfer learning model. During training, the network accuracy is 99.89 percent and the testing accuracy is 84.76 percent. Matlab's converter is used to translate the output (Text to Speech).

***Key Words*:** *Sign Language, Deep Learning, Transfer Learning, ShuffleNet, Neural Network.*

1. INTRODUCTION

The early sign language was a direct result source of using hands to interact. For the language barriers between trades, sign language helped the vendors. Deaf people don't have religious rights and to get married because it was a shameful disability, but a few people broke this. Sign languages (also known as signed languages) use the visual-manual modality to convey meaning. There are approximately 300 sign languages in use around the world today. In 2020, India's population will be 138 crores, with 63 million people suffering from either complete or partial deafness.

2. RELATED WORK

Sign languages (SLs) are different as they use the" corporal - visual" channel, produced with the body and perceived with the eyes Deaf and Dumb people are used Sign Language to communicate with ordinary people. [1] They recognized ASL (American Sign Language) characters using a capsule network (CapsNet). Around the world, a variety of sign languages are used. The American Sign Language (ASL) is a one-of-a-kind sign language (Sign Language).CapsNet has been improved to classify ASL characters with an accuracy of 95.08%. There are only 26 sign letters in ASL. It's so simple to train and classify. [2] In MATLAB, we are performing Sign Language Recognition. CNN, Tensorflow, OpenCV, and Histogram back-projection were used to perform sign language recognition. They achieved an overall accuracy of 80%. [3] In the future, a better dataset and a better network will improve the accuracy of Hand Gesture Recognition. They got 91.37 % of the training accuracy and 87.5% of the testing right. Based on this paper, I'm replacing MobileNetV2 with ShuffleNet and creating a dataset to improve accuracy. [4] The Sign will be detected in a specific area in this study. The use of this method has made cutting the hand sign easier. This method is used to pre-process an input image. [5] They used data augmentation to create a Chinese Sign Language (CSL) benchmark dataset. In data analysis, data augmentation is a technique used to increase the amount of data by adding slightly modified copies of previously existing data or newly created synthetic data from previously existing data. Training a machine learning model acts as a regularized and helps reduce over-fitting. In this research, Transfer Learning is more suitable for high-speed, low latency applications. [6] They achieved a 99.4 % accuracy by combining convolutional neural networks using the Ensemble method. Ensembling combines multiple learning algorithms to obtain their collective performance, i.e., to improve the performance of existing models by combining several models, resulting in a single reliable model. As a result, the Ensemble method can improve the accuracy of an impact. [7] InstantSL is one of the methods for recognizing sign language. In- stantSL proposes a unified approach that includes SL recognition technologies such as machine learning for detection accuracy and voice translation for user-friendliness. The mobile operating system Android for flexibility. [8] Hand tracking and hand representation are two critical components of the proposed Sign Language Recognition system in this work. A model is trained using datasets from RWTH-BOSTON-104, RWTH-BOSTON-50, and ASLLVD. They used a single pre-trained Convolution Neural Network model to recognize Sign Language. To illustrate a hypothetical hand, they used four alternative ways. Gait energy image (GEI), hand energy image (HEI), motion energy image (MEI), and motion history image are some examples (MHI). Finally, the RWTH-BOSTON-50 dataset CNNT method with HEI achieves a higher recognition rate and accuracy of 89.33%. [9] MAST (Myo Armband Sign-Language Translator) is a revolutionary algorithm. That collects muscle electromyography signals using a Myo armband sensor and then classifies them using an upgraded version of a dynamic random forest to translate hand motions into medical sign language. Compared to a popular classification scheme like Support Vector Machines (SVM) and a deep learning technique like Convolutional Neural Network (CNN), the results show an improvement of more than 20%. Compared to Support Vector Machine (SVM), MAST's Random Forest accuracy is 95% compared to Common CNN ResNet-18. Compared to CNN, SVM, and traditional random forest techniques, MAST can recognize different gestures accurately and with high generalization capacity. [10] Sign language is classified into four steps. Image pre-processing for skin detection, component detection (face detection and skin color detection), component localization for hand position, and recognition of Sign Language are all available. Overall, the accuracy is 96.36%. [11] To recognize sign language, they used

the pre-train CNN model VggNet. The Dataset includes 20,000 sign images with ten static digits. The system's accuracy is 97.62%. [13] Their project is divided into two parts. The first component is hand detection, and the second component is signed recognition. The first part employs a Single Shot multi-box Detector (SSD). The second module consists of Convolution Neural Network (CNN) and Fully Connected Network (FCN) to classify a sign language (FCN). Finally, the accuracy of their testing was 92.21%. [16] They develop their own Convolution Neural Network for Sign Language Conversion. Thirty-two thousand images are used to train a CNN model, achieving 97% accuracy. This study used hardware such as a Raspberry Pi Model B, OpenCV, Tensorflow, and Keras.

3. METHODOLOGY

The proposed algorithm and its flow of operation have been organized in this section using a flow diagram and principles, among other things.

3.1. DATASET

There are 19,500 sign language images containing alphabet datasets created using the computer's built-in camera. Each alphabet has 750 images(A-Z), and each image has a memory size of 3 KB. The Dataset for American Sign Language can be viewed in (Figure 2).

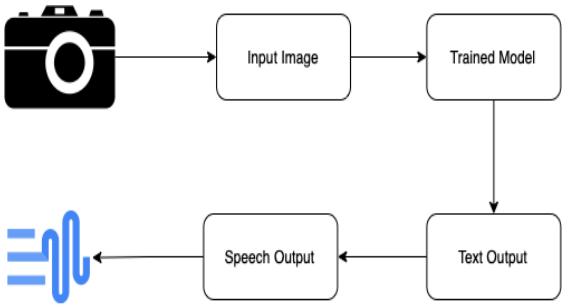


Fig-1: Block diagram of the proposed system

3.2. TRAINED MODEL

The pre-trained network is used to train the dataset and there are approximately 17 pre-trained networks. While comparing it with other pre-trained networks, it gives 96.97%. So, we chose the shufflnet for creating the transfer learning model.

3.3. SHUFFLENET

A Mobile-Friendly Convolutional Neural Network. ShuffleNet is a CNN architecture designed for mobile devices. Three pioneering CNN architectures inspire ShuffleNet's architecture: GoogleNet for its group convolution, ResNet for skip connections, and Xception for its depthwise separable convolutions. ShuffleNet has two types of Blocks. (Figure 5)The basic unit of ShuffleNet V1 and The basic unit of ShuffleNet V1 with stride 2. (Figure 4) ShuffleNet consists of a combination of shuffleNet V1 (Fig-4 BLOCK a) and shuffleNet V1 with stride 2 (Fig-4 BLOCK b) of 18 blocks and totally 172 layers.

3.4. TRANSFER LEARNING

Transfer learning is a popular approach to computer vision. Computer vision training begins with pre-trained models. GoogleNet, DenseNet, ShuffleNet, AlexNet, and other networks are examples. Train your data to create another trained net for image classification.

Fig-2: American Sign Language Characters

3.5. CHANNEL SHUFFLE OPERATION

Before being transmitted into the shuffle channel, the filter was applied to the input image. This filter separates each pixel in an input image. Feature extraction is a step in the dimensionality reduction process that divides and reduces a large set of raw data into more manageable numerical feature groups. It is known as Feature Extraction. A key observation with the grouped pointwise convolution setup is that multiple group convolutions stacked one after the other result in outputs from only specific channels. For example, suppose there are four groups, each with three channels. After pointwise convolution of the first group, the output will consist of representations limited to the first group and not the others. According to the authors of ShuffleNet, this property prevents information flow between channel groups and weakens the model. They introduced the Shuffle operation to address this, which jumbles up the channels across groups. This shuffle does not occur randomly; (Figure 3) illustrates the steps with an example. A different color represents each group to help visualize the shuffling operation.

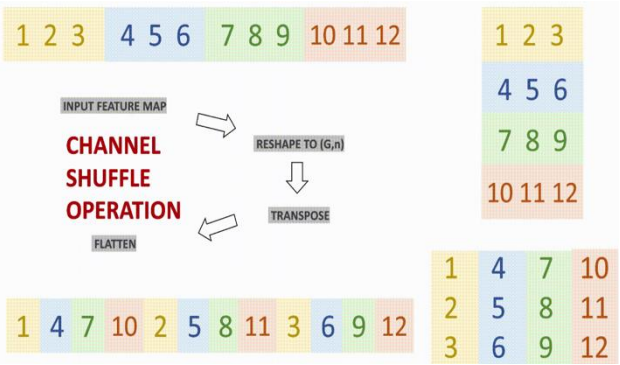


Fig-3: Channel Shuffle Operation

G denotes the number of groups, and n indicates the number of channels within each group. The shuffle operation joins features from different channels, which strengthens representations. Another advantage of channel shuffle operation is its uniqueness.

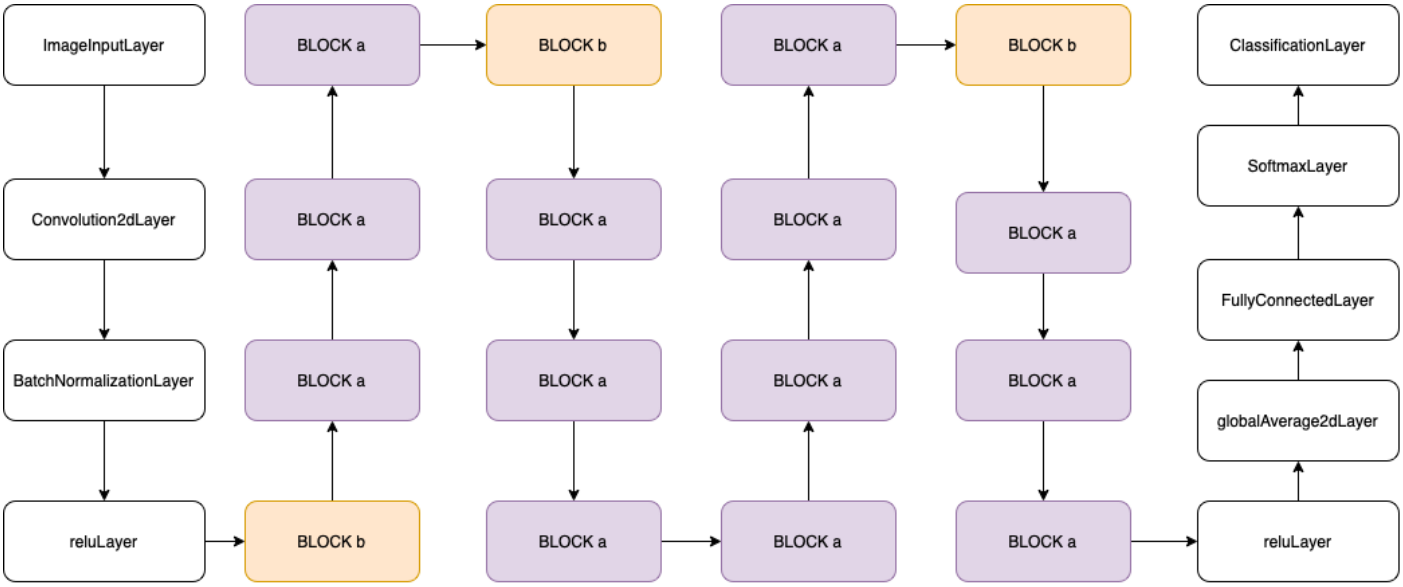


Fig-4: ShuffleNet Block Diagram

The diagram (Figure 5) depicts a typical ShuffleNet in a skip connection network, with pointwise group convolution and channel shuffle operation. Figure 5a has stride=1, while Figure 5b has stride=2 and concatenate operation to support the dimensions of output feature representations. It comprises two pointwise functions, and the second is used to match the dimension to the skip connection. The shuffle operation for the second pointwise convolution is not performed because it produces comparable results when performed. (Figure 5)For stride=2, two modifications are made: 3X3 average pooling is added on the skip connection path to account for spacial activations of feature vectors via average pooling that may be lost in Depth Convolution with stride=2. Concatenation is used instead of simple addition to increase channel dimension with the least amount of computation.

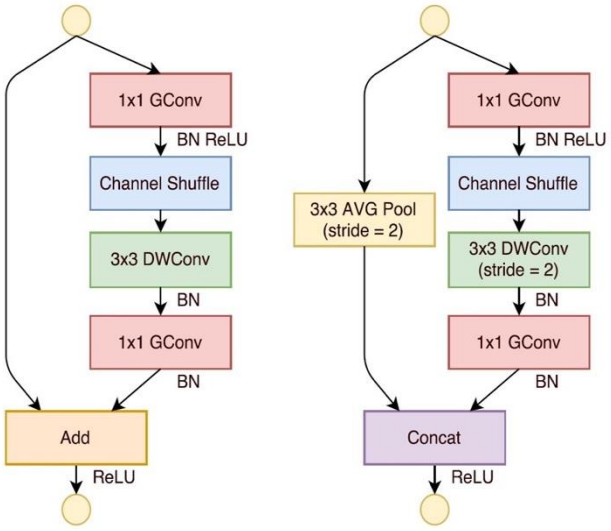


Fig-5: (a) the basic unit of ShuffleNet v1, (b) the basic unit for scaling down in ShuffleNet v1

3.6. TESTING AND VALIDATING THE TRANSFER LEARNING MODEL

After training, the size of a trained model is 20 MB for 19,500 images. Firstly the saved Net is loaded, and the unknown Hand Sign is given to classify the text. Further, the text was converted to Speech. The final output of this project is a Speech format. Text can be converted to Speech using MATLAB from the outcome of a trained model. It is possible to obtain it using two MATLAB functions. As a result, the system() and sprint() functions convert text to Speech.

4. RESULTS AND DISCUSSION

A total of 19,500 images are used to train a transfer learning model. Were the 26 folders contain an equal number of images. The number of images also increases the accuracy simultaneously increases. A training output progress graph is shown in Figure 6. It displays a training output during the transfer learning training of a pre-trained model. The training output table shows a Mini-batch accuracy, Base Learning Rate, Mini-batch Loss, and time taken to complete a dependent Iteration are all shown in the training output table 1. Finally, I achieved 99.44 percent training accuracy.

Table-1: Training Parameters and it's Accuracy.

Epoch	Iteration	Time Elapsed in Minutes	Mini-batch Accuracy	Mini-batch Loss	Base Learning Rate
1	1	00:00:04	1.56%	15.1293	0.1
3	50	00:01:42	99.44%	0.0886	0.1
5	100	00:03:20	99.78%	0.0354	0.1
5	105	00:03:29	99.89%	0.0177	0.1

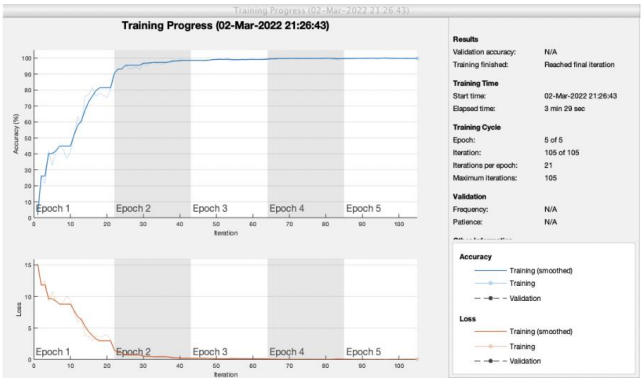
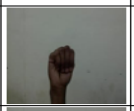
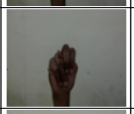
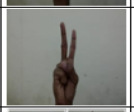


Fig-6: Training Progress of the proposed system

5. CONCLUSION AND FUTURE SCOPE:

In this paper, the transfer learning model is used recognize a sign character. The testing accuracy and time taken for a trained model to recognize and classify a Sign using input images. In the trained model, testing accuracy was 84.76 percent. It plans to create a hand-held device for converting sign language to Speech in the future. Furthermore, the device can be developed using an AR glass and a hearing machine. AR glasses are used to reduce the compatibility of carrying that device. Hearing machine is helpful for a deaf person who is incompatible. The research work can be implemented on the computer like normal people use the Hello google command. They can interact and search for their needs. Finally, we used Matlab to simulate a Sign Language to Speech conversion.

Table-2: Prediction of Characters and Sentences from images

S.No.	Input Image	Input Label	Model Predicted Label	Time Taken for Validation
1		M	N	Elapsed time is 0.033804 seconds.
2		N	N	Elapsed time is 0.034226 seconds.
3		V	V	Elapsed time is 0.029578 seconds.
4		Nice to meet you	Nice to meet you	Elapsed time is 6.433451 seconds.
5		what is your name	what is your name	Elapsed time is 1.203948 seconds.

6. REFERENCES

1. Metin Bilgin, Korhan Mutludoğan "American Sign Language Character Recognition with capsule Network", 2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) (2019).
2. Prosenjit Roy, S.M Miraj Uddin, Md. Arifur Rahman, Md. Musfiqur Rahman, Md. Shahin Alam, Md. Saidur Rashid Mahin "Bangla Sign Language Conversation Interpreter Using Image Processing", 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT) (2019).
3. Shruti Chavan, Xinrui Yu, Jafar Saniie "Convolution Neural Network Hand Gesture Recognition for American Sign Language", 2021 IEEE International Conference on Electro Information Technology (EIT) (2021).
4. Bayan Mohammed Saleh, Reem Ibrahim Al-Beshr, Muhammad Usman Tariq "D-Talk Sign Language Recognition System for people with Disability using Machine Learning and Image Processing" International Journal of Advanced Trends in Computer Science and Engineering · September 2020.
5. Yuzhe Ding, Shaofei Huang, Roubo Peng "Data augmentation and Deep Learning Modeling Methods on Edge-Device-Based Sign Language Recognition", 2020 5th International Conference on Information Science, Computer Technology and Transportation (ISCTT) (2020).
6. Citra Suardi, Anik Nur Handayani, Rosa Andrie Asmara, Aji Prasetya Wibawa, Lilis Nur Hayati, Huzain Azis" Design of Sign Language Recognition Using E-CNN", 2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT) (2021).
7. Deshinta Arrova Dewi" InstantSL: A Sign Language Model to Support Two-Way Communication Between Aurally Impaired Communities with Others", 2019 International Journal of Innovative Technology and Exploring Engineering (IJITEE) (2019).
8. Kian Ming Lim, Alan Wee Chiat Tan, Chin Poo Lee & Shing Chiang Tan "Isolated Sign Language Recognition using Convolution Neural Network Hand Modeling and Hand Energy Image" Multimedia Tools and Applications volume 78, pages 19917–19944 (2019).
9. Zuhaib Muhammad Shakeel, Soonhyuk So, Patrick Lingga, Jaehoon Paul Jeong" MAST: Myo Armband Sign Language Translator for Human Hand Activity Classification", 2020 International Conference on Information and Communication Technology Convergence (ICTC) (2020).
10. Sai Myo Htet, Bawin Aye, Myo Min Hein" Myanmar Sign Language Classification using Deep Learning", 2020 International Conference on Advanced Information Technologies (ICAIT) (2020).
11. Karma Wangchuk, Panomkhawn Riyamongkol, Rattapoom Waranusast "Real-time Bhutanese Sign Language digits recognition System Using Convolution Neural Network", ICT Express Volume 7, Issue 2, June 2021, Pages 215-220 (2021).
12. Müjde Aktaş, Berk Gökberk, Lale Akarun "Recognise Non-Manual Signs in Turkish Sign Language", 2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA) (2019).
13. Rahib Abiyev, John Bush Idoko, Murat Arslan" Reconstruction of Convolution Neural Network for Sign Language Recognition", 2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE) (2020).
14. Boris Mocialov, Graham Turner, Helen Hastie" Transfer Learning for British Sign Language Modelling", Proceedings of the Sixth Workshop on NLP for Similar Languages, Varieties and Dialects, VarDial (2018).
15. Devjoyti Aich, Abdulla Al Zubair, K. M. Zubair Hasan, Antora Deb Nath, Zahid Hasan "A Deep Learning Approach for Recognizing Bengali Character Sign Language", 2020 11th International Conference on Computing,

Communication and Networking Technologies (ICCCNT) (2020).

16. Lakshmi Boppana, Rasheed Ahamed, Harshali Rane, Ravi Kishore Kodali" Assistive Sign Language Converter for Deaf and Dumb" 2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) (2019).