Journal of Theoretical and Applied Information Technology
31st August 2023. Vol.101. No 16
© 2023 Little Lion Scientific

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MULTI-USER REAL-TIME SIGN LANGUAGE RECOGNITION OF WORDS USING TRANSFER LEARNING OF DEEP LEARNING NEURAL NETWORKS

RASHMI GAIKWAD[1], LALITA ADMUTHE[2]

[1]Research Scholar, DKTE Society's Textile & Engineering Institute, Ichalkaranji, Maharashtra, India.

[2]Professor, DKTE Society's Textile & Engineering Institute, Ichalkaranji, Maharashtra, India.

E-mail: [1]rsgaikwad2020@gmail.com, [2]ladmuthe@gmail.com

ABSTRACT

The hearing and speech impaired people use sign language for communication. Nevertheless, other people cannot understand sign language. A real-time American Sign Language recognition system will help to reduce this gap of communication. This paper presents a solution for real-time sign language recognition of words in American Sign Language. In this research transfer learning of two deep learning pre-trained modules available in the Tensorflow object detection repository namely SSD MobileNet V2 FPNLite 320x320 and SSD ResNet50 V1 FPN 640x640 is implemented. The dataset consisting of signs of eight words is generated exclusively for this research. Signs obtained from a single user are used for training and testing of the networks but real-time detection can be done on signs performed by multiple users. A comparison of performance of the two networks is also done for the same dataset. Accuracy in terms of Confidence level is 100% for same signer detection and for different signers it comes in between 60% to 80%.The precision and recall of SSD MobileNet V2 came to be 91% and 71% respectively while that of SSD ResNet50 V1 came to be 87% and 74% respectively.

Keywords: *Deep Learning Neural Networks, Sign language recognition (SLR), SSD MobileNet V2 FPNLite 320x320, SSD ResNet50 V1 FPN 640x640, Convolutional Neural Network (CNN), Dataset.*

1. INTRODUCTION

Sign Language is a way of communication between the people with hearing and speech impairment and normal people. Normal people don't understand sign language very well. So, an intelligent sign language recognition system which requires minimum efforts from the signer to perform signs will be useful in bridging the communication gap between the two communities.

Sign language recognition can be divided into two main categories. First is the glove based approach where the signer is required to wear gloves mounted with sensors and data will be collected and processed from the sensors. Second is the computer vision based approach. In this approach, object detection, is one of the most crucial task which deals with location and identification of targets. Deep learning algorithms have been used by a number of research scholars to do object detection in recent years. One such example is recognition of signs. Computer vision based approach is subdivided into static and dynamic sign language recognition systems. Static signs are processed and detected from images of signs which may be captured in real time or from an available dataset. Dynamic signs are captured in real time by a web camera and then processed and detected. The sign language recognition part is carried out by an artificial intelligence (AI) network. The AI network used must be flexible, accurate and precise. The accuracy of the AI network must not be affected by the background conditions and variations.

The people with hearing impairment use various sign languages for communication. Some of them are American Sign Language (ASL), Chinese Sign Language, Indian Sign Language (ISL), Arabic Sign Language etc. Indian sign language varies not only from one state to another state but also from region to region. Comparatively ASL is easier to do research work as it is less complicated and is not varying like ISL. So, this work is focusing on ASL but it can be applied on ISL also.

2. RELATED WORK

Over the year's sign language recognition systems is done using deep learning neural networks. Convolutional Neural Network based sign language recognition systems are most widely used because of their higher accuracy and precision. One such CNN model for ASL recognition is used in [1]. This system is designed to recognize signs of letters only and is implemented using Keras, Tensorflow and openCV. The accuracy of recognition is 99.838%. Same CNN model is used for the recognition of Bhutanese sign language of digits in [2]. They have also compared the results of CNN with SVM, KNN, Logistic regression and LeNet5. The performance was evaluated using the parameters like accuracy, precision, recall and F1-score.

Generating text and speech from the input gesture sign language is achieved in [3]. In [4] sign language recognition is done using inflated 3D convolutional networks i.e. I3D ConvNet. Here, the ChaLearn249 dataset is used which consists of images and videos of sign language. In [5] deep learning using CNN is used to locate the signer hands by using 5 layers of CNN and average pooling. The activation function used is ReLu. Long and short term memory neural network (LSTM) is used for encoding and decoding of input frames of variable length by obtaining the temporal structure information. The accuracy of the recognition rate of the network used in this paper is 99%.

A 3D combinational neural network sign language recognition system by extracting spatial-temporal information of signs is proposed in [6]. The region of interest i.e. ROI is the part in an image frame which consists the information about the gestures. All the background is subtracted from the image frame which is obtained from the signer video. The work is focused on large vocabulary Chinese sign language recognition. This system is only implemented for static images and only the signs of words are detected. An accuracy of 94.3% is achieved using this technique. A real time traffic sign recognition system using faster recurrent combinational neural networks and MobileNet is proposed in [7]. The limitations of colors and space which vary from sign to sign have been eliminated by using the RGBN color space and detection of contours and centers of traffic signals. An accuracy of 84.5% and recall of 94% is achieved using this method. A fully convolutional neural network for the detection and recognition of traffic signs is used in [8]. This work is divided into two stages. In the first stage the size and orientation of the traffic sign is detected and in the second stage the text of the traffic sign is detected. A precision of 93.5% and recall of 94% is achieved using this technique.

A method to detect signs of English letters from A to Z in American Sign Language by using Neural Networks is proposed in [9]. The results in this paper show that the speed of processing is improved by the use of neural network and the dataset can be increased to any number of variables. A comparison of the efficiency of different types of Neural Networks in Arabic sign Language Gesture recognition for static as well as dynamic signs is done in [10] and proved that the fully recurrent neural networks have the highest accuracy and minimum error rate. A database of signs for 28 letters of Arabic Sign Language is generated and after training the Neural Network and the signs are detected. The accuracy of recognition using this technique is found to be 95.11%.

Some researchers have used hardware like leap motion controllers[11,16], Kinect sensors[13,14], accelerometer and gyroscopic sensors [25,26] and other such hardware devices [26] to capture hand movements while performing signs and after feature extraction the dataset is fed for training of the AI network which may be a CNN, Support vector machine[16,22], Self-organizing maps[20], Dynamic time warping algorithm [13], fuzzy networks[17,18,19], Transition movement models[21] and Principal Component Analysis[30,31]. A method to recognize Indian sign language gestures with the use of flex sensors, gyroscope, accelerometer and microcontroller is proposed in [12]. A platform for hardware and software i.e. microcontroller programming and interfacing is achieved by making the use of Arduino.

Signs of words in a video are recognized using Reinforcement learning and spatial-temporal CNN and bidirectional LSTM in [23].The performance of this system is mapped using the parameter word error rate which comes 28.6% in this research. Hidden Markov Model and K-nearest neighbor are used for recognition of signs in [24]. Real time Indian sign language recognition by using image processing techniques is done in [27]. A system for dynamic sign language recognition system for smart home interaction is proposed in [28]. Meaningful sentences are formed by the use of stochastic linear formal grammar (SLFG) module. This system is not applied to real-time videos and it can be expanded for sign language communication in real time. The accuracy with this system is 98.65%.

Convolutional neural network in combination with long short-term memory (LSTM) in [29] is the most recent development in sign language recognition where the data is collected and signs are recognized using OpenCV and computer vision. This method achieved an accuracy of 95.5%. A human posture recognition system for video surveillance using different classifiers like K Means, Fuzzy C Means, and Multilayer Perceptron, Self-organizing Maps and Neural Networks is proposed in [32]. In [33], a combination of ANN, k-nearest neighbor and support vector machine is used for Italian sign language recognition. This system achieves an accuracy of 93.1%. In [34], dynamic hand gesture recognition system is implemented by using a fusion of 1-D and 2-D CNN with temporal convolutional network. 16 features of each sample are generated using data gloves. An attention-based encoder-decoder model with a multi-channel convolutional neural network is proposed in [35] with a Word Error Recognition (WER) = 10.8%. It translates Chinese sign language into voices.

3. METHODOLOGY

Fig. 1 shows the steps of the research work carried out. The data acquisition is done by recording signs performed by the signer in front of the web camera. These signs are fed to the Neural Networks for transfer learning after pre-processing and then after training and testing of the network, real-time sign detection is done.

Deep learning neural network consists of convolutional layer, MaxPooling layer and fully connected layer as shown in fig. 2. It is a very tough task to build a computer vision model from scratch because to make the model generalize well a wide variety of input data is needed, and training such models can take several hours or days on a GPU. So, for an easier and faster model we are using transfer learning. This way, only training the upper layers of a neural network of the well trained model leads to much more reliable results.

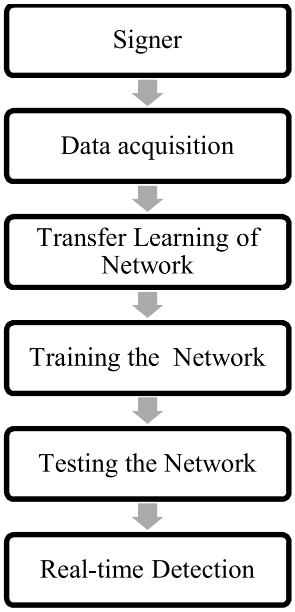


Figure 1: Flow Of Research Work

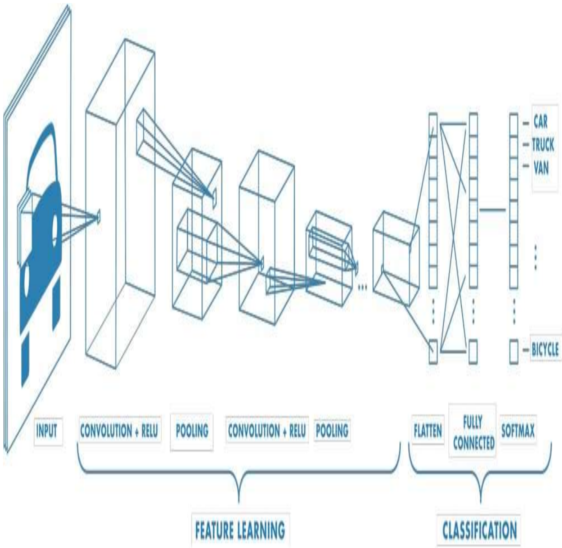


Figure 2: A Deep Learning Neural Network [38]

Tensorflow Object detection API is the framework for creating a deep learning network that solves object detection problems. This module consists of a variety of detection models which are pre-trained on the COCO 2017 dataset. In this research SSD MobileNet V2 FPNLite 320x320 and SSD ResNet50 V1 FPN 640x640 are implemented for the detection of signs of words in real-time. Transfer learning is used to train the two models on the dataset generated in this research. The architecture of the two deep learning models can be divided into three parts a base or backbone network, a feature extractor and a detection network as shown in fig. 3.

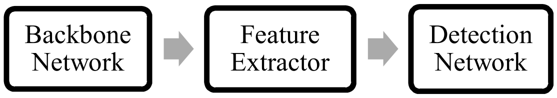


Figure 3: Architecture Block Diagram of Deep Learning Models

The backbone network that is MobileNetV2 and ResNet50, are nothing but neural networks. If we use a fully connected layer and a softmax layer at the end of these networks, we get a classification network. For detection purpose, these layers are replaced with detection networks like SSD to perform object detection i.e. in this case the detection of signs. The feature extractor is called as Feature Pyramid Network (FPN) as shown in fig. 4 is implemented to improve the accuracy and speed. The FPN makes the shallow feature map have additional semantic information. Object detection is performed by the SSD using multiple feature maps. As shallow feature maps possess detail information but suffer from a shortage of semantic information therefore, the inclusion of the FPN structure in SSD can successfully overcome this drawback and enable the SSD to detect small objects.

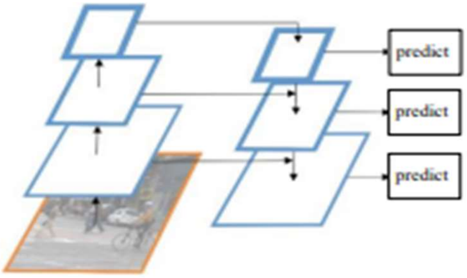


Figure 4: A Feature Pyramid Network [36]

The detection network is the SSD (Single Shot Detector) architecture which is a single convolution network. It learns to predict bounding box locations and categorize these locations in one pass. If regional proposal networks (RPN) based approaches are used such as R-CNN series then they need two shots, one for generating region proposals and one for detecting the object of each proposal. So, SSD is much faster compared with other detection networks. SSD performance might get degraded if employed for detecting objects that are too close or too small so the FPN network is included before SSD as stated above.

Following is the step by step procedure divided into 4 stages which are- data acquisition, training the network, testing the network and sign recognition in real-time.

3.1 Data Acquisition

A dataset of signs of eight words 'hello', 'livelong' 'yes', 'no', 'thank you' 'thumbs up', 'thumbs down' and 'one' was created using openCV. Images were captured using web camera. 100 images of each word were captured so the dataset consisted of 100x8 = 800 images for training from a single user and testing dataset consisted of 15x8 = 120 images. Labeling of images was done using the graphical image annotation tool labelImg where only the hand part was annotated.

3.2 Training the Network

The mathematical expressions describing the training of the SSD [37] are as follows. Let x_{ij}^{p} = {1, 0} indicate the ith default box matching to the jth ground truth box of category p. Thus, we have,

$$\sum_i x_{ij}^{p} > 1 \qquad (1)$$

The weighted sum of the localization loss (locl) and the confidence loss (confl) is the total loss function L given as,

$$L(x,c,l,g) = \frac{1}{N}\left(L_{confl}(x,c) + \alpha L_{locl}(x,l,g)\right) \qquad (2)$$

where N is the number default boxes matched and the localization loss is the loss between the parameters of the predicted box (l) and the ground truth box (g). The confidence loss is the softmax loss over multiple class confidences (c).The weight term α is set to1 by cross validation.

Consider that there m feature maps for prediction. For each feature map the scale of the default boxes is calculated as,

$$s_k = s_{min} + \frac{s_{max} - s_{min}}{m-1}(k-1), \;\; k \,\epsilon\, [1,m] \qquad (3)$$

The width and height for each default box is given as,

$$w_k^a = s_k\sqrt{a_r} \qquad (4)$$

$$h_k^a = s_k/\sqrt{a_r} \qquad (5)$$

In this research, first the SSD MobileNet V2 network is trained on 100 images of each sign by transfer learning of the pre-trained model and tested on 15 images of each sign. The number of steps for training (epoch) is set to 2000 to achieve best results. The learning rate goes on increasing till 0.08 and then it becomes constant as shown in fig. 5. The loss metric while training the network starts at 0.7 and goes on reducing till approximately 0.4 at the last step of training as shown in fig. 6.

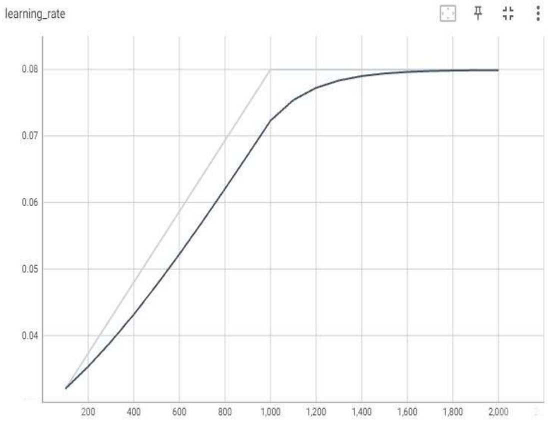


Figure 5: Learning Rate of SSD MobileNet V2

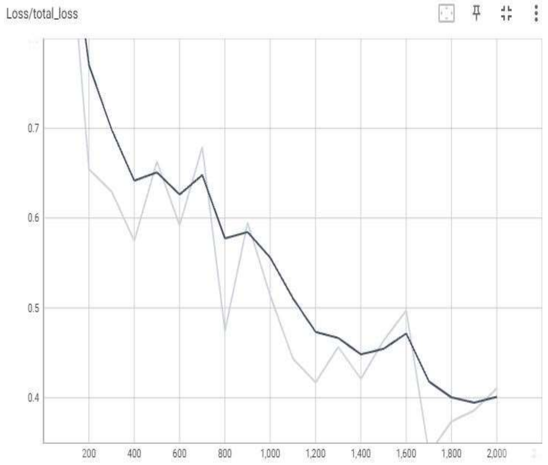


Figure 6: Total Loss of SSD MobileNet V2

The same dataset is used for training and testing of the second network i.e. SSD ResNet50 V1. The learning rate of this network goes on increasing till 0.04 as shown in fig. 7. The loss metric while training the network goes on reducing till 0.5 as shown in fig. 8. This training of networks was done on the computer without GPU.

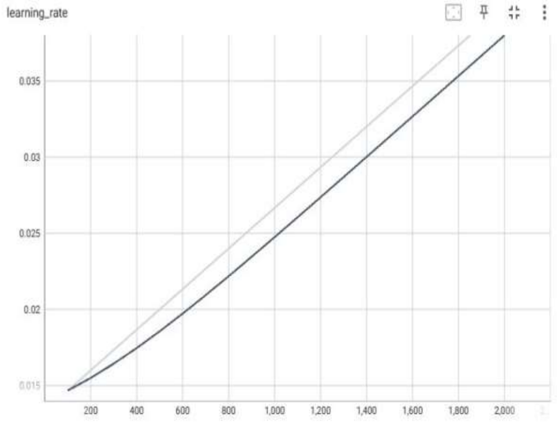


Figure 7: Learning Rate of SSD Resnet50 V2

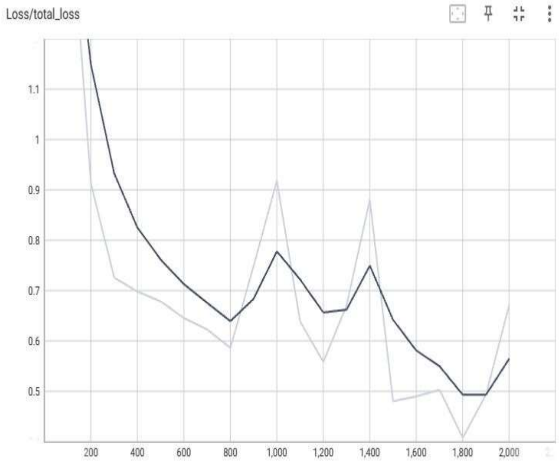


Figure 8: Total Loss of SSD Resnet50 V2

3.3 Testing the Network

The trained network is tested on 120 different images of the same signs. In this research, Average Precision (AP) and Recall (R) are the evaluation parameters and accuracy in terms of confidence level in percentage is used to indicate the detection of signs in real-time. The mAP indicates the mean of the average precision (AP) of the classes as shown in equation (6) and the number of classes is denoted by N(C). AP is determined by recall (R) and precision (P) as shown in equation (8) & (9) respectively. FP and TP means the amount of False Positive and True Positive. FN means the amount of False Negative.

$$mAP = \frac{\sum AP}{N(C)} \qquad (6)$$

$$AP = \int_0^1 P(R)dR \qquad (7)$$

$$R = \frac{TP}{FN+TP} \qquad (8)$$

$$P = \frac{TP}{FP+TP} \tag{9}$$

The maximum precision and recall values of SSD MobileNet V2 are shown in fig. 9 and fig. 10 respectively and that of SSD ResNet50 V1 is shown in fig. 11 and fig. 12 respectively. The values are also tabulated along with the time required to train the two networks in table 1. The time taken for MobileNet V2 is 1-2 hours whereas that for SSD ResNet50 V1 is nearly 24 hours.

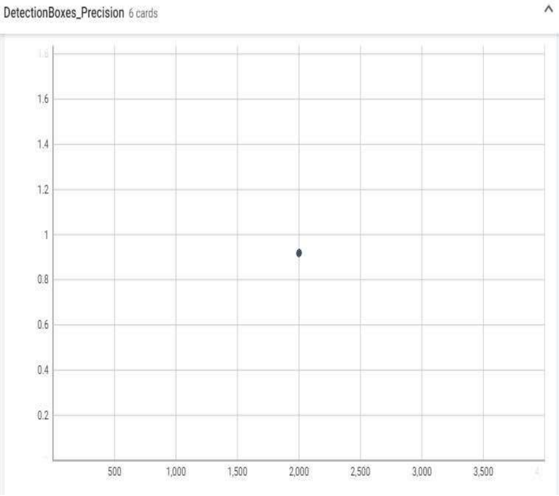


Figure 9: SSD MobileNet V2 Precision

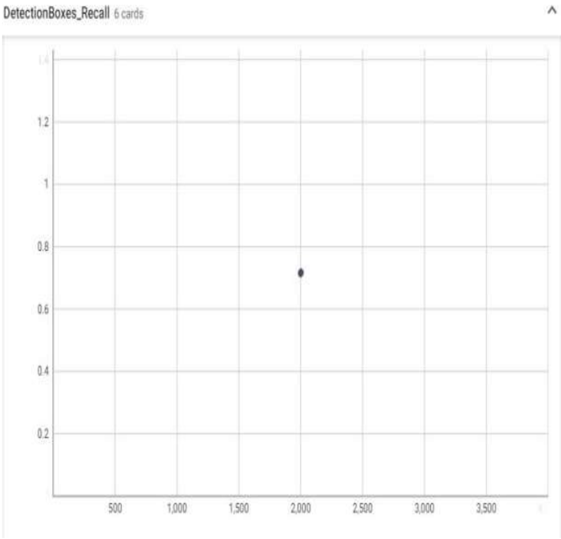


Figure 10: SSD MobileNet V2 Recall

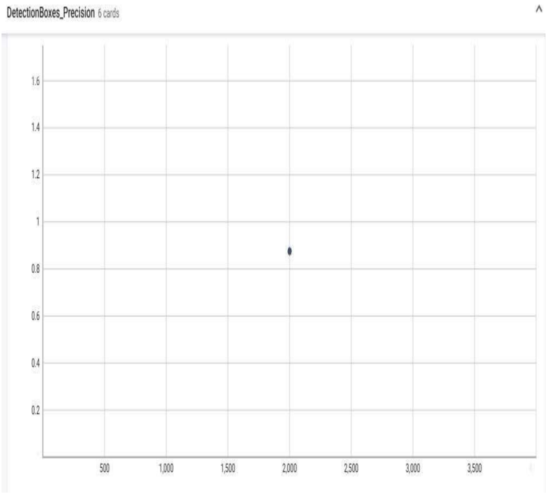
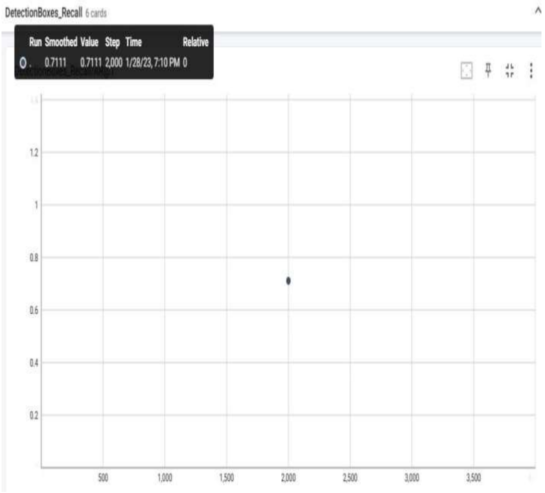


Figure 11: SSD ResNet50 V1 Precision



Run	Smoothed	Value	Step	Time	Relative
○ .	0.7111	0.7111	2,000	1/28/23, 7:10 PM	0

Figure 12: SSD ResNet50 V1 Recall

Table 1*: Comparison of Training and Evaluation Parameters of the Two Networks.*

3.4 Sign Detection in Real-time

Once the network is trained and tested on the dataset of the signs the model can be used to detect real-time signs performed by a person in front of the web camera. SSD uses bounding boxes for every image category while detecting signs in real time. Fig. 13 and fig. 14 show that the confidence level of real- time detection of the sign "Hello" is 100% using both the networks respectively.

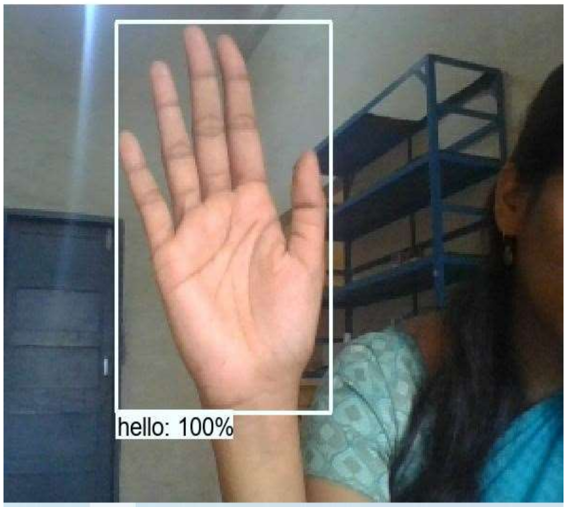


Figure 13: Hello sign detected in real-time using SSD MobileNet V2

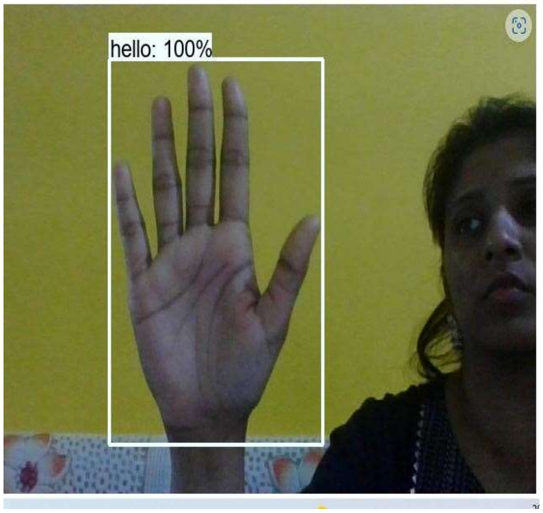


Figure 14: Hello sign detected in real-time using SSD ResNet50 V1

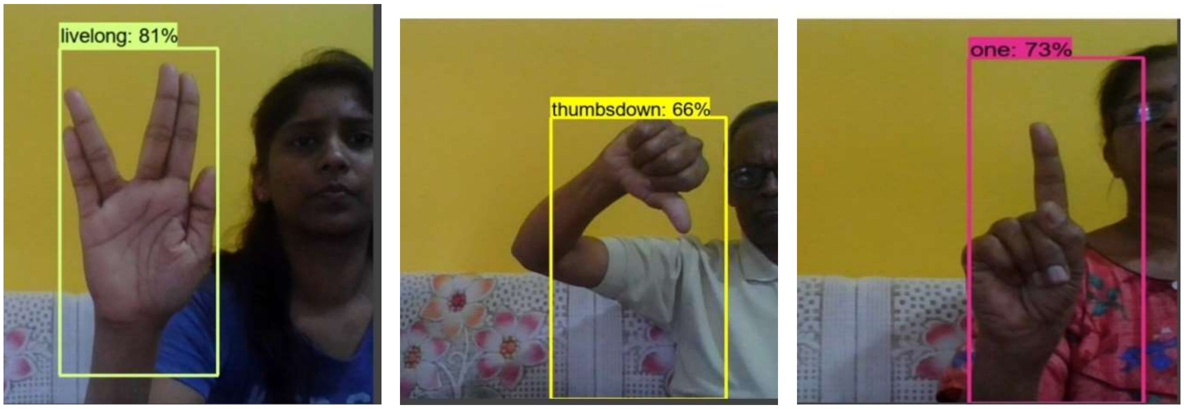
4. RESULTS

After training and testing of the SSD MobileNet V2 network real-time signs of the words are detected successfully through the webcam. The same process of training and testing is done for another network SSD ResNet50 V1. It is observed that the time required to train SSD MobileNet V2 is much less (nearly 1 hour) than that required for SSD ResNet50 V1 (nearly 24hours). This training is without any graphics card available on the machine. It is also observed that the time for training and the confidence level of the signs detected in real time goes on increasing if the number of input images is increased from 50 to 100 images per sign.

Network	SSD MobileNet V2 FPNLite 320x320	SSD ResNet50 V1 FPN 640x640
Learning Rate at epoch 2000	0.08	0.04
Total Loss at epoch 2000	0.41	0.5
Highest Precision Value Achieved	0.919	0.876
Highest Recall Value Achieved	0.716	0.743
Time Required for training of network on computer without GPU	1-2 hours	24hours

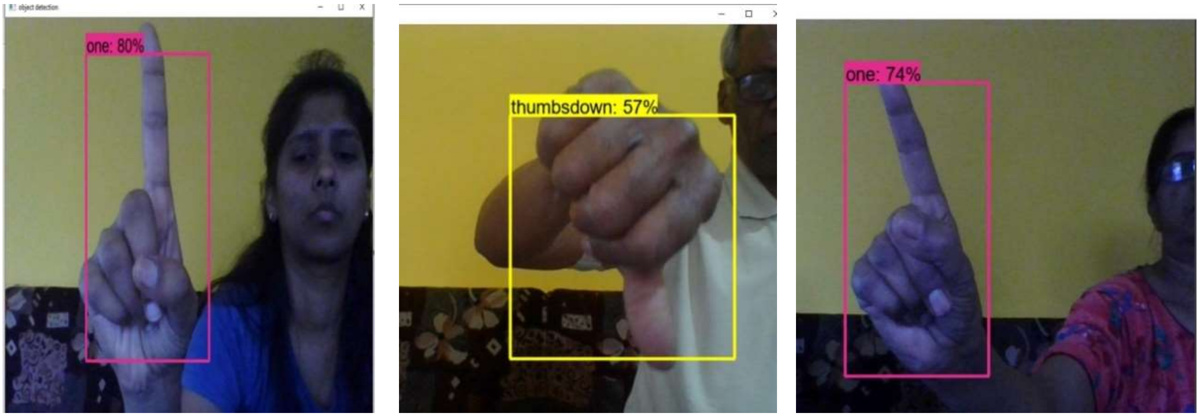
The precision and recall of SSD MobileNet V2 came to be 91% and 71% respectively while that of SSD ResNet50 V1 came to be 87% and 74% respectively. So, there is a difference of around 10% in the evaluation parameters.

Fig. 15 and fig. 16 shows the real-time output of the two networks where three different signers are performing the signs in front of web camera. The accuracy is 65 % to 82% for different signs using SSD MobileNet V2 whereas that for SSD ResNet50 V1 it is 56% to 80%. After comparing both the networks it is clear that SSD MobileNet V2 is a better choice for multi-user real-time sign language recognition of words.



(a) Signer 1 (b) Signer 2 (c) Signer 3

Figure 15: Real-Time Output of SSD MobileNet V2: (a) Signer 1 (b) Signer 2 (c) Signer 3



(a) Signer 1 (b) Signer 2 (a) Signer 3

Figure 16: Real-Time Output of SSD ResNet50 V1: (a) Signer 1 (b) Signer 2 (c) Signer 3

5. CONCLUSION

In this research, signs of eight words in American Sign Language are recognized in real-time using transfer learning process on two pre-trained deep learning neural networks. OpenCV is used to generate a dataset of signs of 8 words. It consists of total 800 images i.e. 100 images per sign. The eight words are ‘hello’, ‘livelong’, ‘yes’, ‘no’, ‘thank you’ ‘thumbs up’, ‘thumbs down’ and ‘one’. The performance of the two networks SSD MobileNet V2 FPNLite 320x320 and SSD ResNet50 V1 FPN 640x640 is compared using evaluation parameters like confidence level, precision, recall and time required for training of the network. Accuracy in terms of confidence level is 100% for same signer detection and for different signers it comes in between 60% to 80%. The precision and recall of SSD MobileNet V2 came to be 91% and 71% respectively while that of SSD ResNet50 V1 came to be 87% and 74% respectively. In future work, more signs of words can be included in the dataset. In order to improve accuracy of sign recognition signs obtained from multiple users can be included. Generation of sentences from the detected words using grammar rules can be done in future based on this research.

REFERENCES:

[1] A. Kasapbasi, A. Eltayeb A, A. Elbushra, O. Al-Hardanee, A. Yilmaz, “DeepASLR: A CNN based human computer interface for American Sign Language recognition for hearing impaired individuals”, Elsevier, 2022, ISSN: 2666-9900.

[2] K. Wangchuk, P. Riyamongkol, R. Waranusast, “Real-time Bhutanese Sign Language digits recognition system using convolutional neural network”, 2020, ISSN:2405-9595, The Korean Institute of Communications and Information Sciences(KCIS), Elsevier B.V.

[3] A. Thakur, P. Budhathoki, S. Upreti, S. Shrestha, S. Shakya, “Real-time Sign Language Recognition and Speech Generation”, Journal of Innovative Image Processing, 2020, Vol.02/ no. 02 pp. 65-76, ISSN: 2582-4252.

[4] N. Sarhan N. and S. Frintrop, “Transfer Learning for Videos: From Action Recognition to Sign Language Recognition”, IEEE ICIP, 2020.

[5] Siming He, “Research of a Sign Language Translation System Based on Deep Learning”, IEEE International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM), 2019.

[6] J. Huang, W. Zhou, H. Li and W. Li, “Attention based 3D CNNs for Large Vocabulary Sign Language Recognition”, IEEE Transactions on Circuits and Systems on Video Technology, 2018.

[7] J. Li and Z. Wang, “Real-Time Traffic Sign Recognition Based on Efficient CNNs in the wild”, IEEE Transactions on Intelligent Transportation Systems, 2018.

[8] Y. Zhu, M. Liao, M. Yang and W. Liu, “Cascaded Segmentation-Detection Networks for Text-Based Traffic Sign Detection”, IEEE Transactions on Intelligent Transportation Systems, 2018, Vol. 19, No.1.

[9] P. Mekala, Y. Gao, J. Fan and A. Davari, “Real-time Sign Language Recognition based on Neural Network Architecture”, 978-1-4244-9592-4/11/©2011 IEEE.

[10] M. Maraqa and R. Abu-Zaiter, “Recognition of Arabic sign language (ArSL) using recurrent neural networks”, Applications of Digital Information and Web Technologies, 2018, pp: 478 – 481.

[11] D. Avola, M. Bernardi, L. Cinque, G. L. Foresti G. L. and C. Massaroni, “Exploiting Recurrent Neural Networks and Leap Motion Controller for the recognition of Sign Language and Semaphoric Hand Gestures”, IEEE Transactions on Multimedia, 2018.

[12] G. Plouffe and A. Cretu A, “Static and Dynamic Hand Gesture Recognition in Depth Data Using Dynamic Time Warping”, IEEE Transactions on Instrumentation and Measurement, 2016 Vol. 65, Issue: 2, pp305 – 316.

[13] S. Y. Heera S, M. K. Murthy, V. S. Sravanti and S. Salvi, “Talking Hands – An Indian Sign Language to Speech Translating Gloves”, International Conference on Innovative Mechanisms for Industry Applications, 2017, 978-1-5090-5960-7/17 IEEE.

[14] Y. Yao and Y. Fu, “Contour model-based hand-gesture recognition using the Kinect sensor”, IEEE Transactions on Circuits Systems and Video Technology, 2014, vol. 24, no. 11, pp. 1935–1944.

[15] Z. Ren, J. Yuan, J. Meng, and Z. Zhang, “Robust Part-Based Hand Gesture Recognition Using Kinect Sensor”, IEEE transactions on multimedia, 2013, vol. 15, no. 5.

[16] P. Saini, R. Saini, S. Behera, D. Dogra and P. P. Roy, “Real-Time Recognition of Sign Language Gestures and Air-Writing using Leap Motion”, Fifteenth IAPR International Conference on Machine Vision Applications (MVA), May 2017, Nagoya University, Nagoya, Japan.

[17] G. Jia, H. Lam, S. Ma, Z. Yang, Y. Xu, B. Xiao, “Classification of Electromyographic Hand Gesture Signals Using Modified Fuzzy C-Means Clustering and Two-Step Machine Learning Approach”, IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2020, Volume28, Issue6.

[18] A. R. Várkonyi-Kóczy and B. Tusor, “Human–computer interaction for smart environment applications using fuzzy hand posture and gesture models”, IEEE Transactions on Instrumentation and Measurement, 2011, vol. 60, no. 5, pp. 1505–1514.

[19] P. R. Futane and R. V. Dharaskar, “Video Gestures Identification and Recognition Using Fourier Descriptor and General Fuzzy Minmax Neural Network for Subset of Indian Sign Language”, 2012, 978-1-4673-5116-4/12, IEEE.

[20] I. Infantino I, R. Rizzo and S. Gagli, "A Framework for Sign Language Sentence Recognition by Common-sense Context", IEEE transactions on systems, man, and cybernetics—part c: applications and reviews, September 2007, vol. 37, no. 5.

[21] G. Fang, W. Gao and D. Zhao, "Large-vocabulary continuous sign language recognition based on transition-movement models", IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 2007, vol. 37, no. 1, pp. 1–9.

[22] Y. Quan,"Chinese Sign Language Recognition Based on Video Sequence Appearance Modelling", IEEE Conference on Industrial Electronics and Applications (ICIEA), 2010, pp1537 – 1542.

[23] C. Wei, J. Zhao, W. Zhou, H. Li, "Semantic Boundary Detection with Reinforcement Learning for Continuous Sign Language Recognition", IEEE Transactions on Circuits and Systems for Video Technology, 2021, Volume31, Issue:3.

[24] M. Hassan, A. K. Khaled and T. Shanableh, "User-dependent Sign Language Recognition Using Motion Detection", International Conference on Computational Science and Computational Intelligence, 2016, IEEE.

[25] X. Yang, X. Chen, X. Cao, S. Wei and X. Zhang X, "Chinese Sign Language Recognition Based on An Optimized Tree", 2016, IEEE.

[26] V. E. Kosmidou and L. J. Hadjileontiadis, "Sign Language Recognition Using Intrinsic-Mode Sample Entropy on sEMG and Accelerometer Data", IEEE transactions on biomedical engineering, 2009, vol. 56, no. 12.

[27] P. Subha Rajam and G. Balakrishnan G., "Real Time Indian Sign Language Recognition System to aid Deaf-dumb People", 978-1-61284-307-0/11/ ©2011 IEEE pp 737-742.

[28] M. R. Abid, E. M. Petriu and E. Amjadian, "Dynamic sign language recognition for smart home interactive application using stochastic linear formal grammar", 2015, IEEE Transactions on Instrumentation and Measurement, vol. 64, no. 3, pp. 596–605.

[29] W. Li, H. Pu, R. Wang, "Sign Language Recognition Based on Computer Vision, 2021, IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA).

[30] M. Ahuja and A. Singh, "Static Vision Based Hand Gesture Recognition Using Principal Component Analysis", IEEE 2015.

[31] M. Mohandes, M. Deriche and J. Liu, "Image-Based and Sensor-Based Approaches to Arabic Sign Language Recognition", IEEE transactions on human-machine systems, 2014, pp 2168-2291.

[32] K. K. Htike, O. O. Khalifa, H. A. Ramli and A. M. Abushariah, "Human Activity Recognition for Video Surveillance using Sequences of Postures", 2014, IEEE ISBN: 978-1-4799-3166-8.

[33] A. Calado, V. Errico, G. Saggio, "Toward the Minimum Number of Wearables to Recognize Signer-Independent Italian Sign Language with Machine-Learning Algorithms", IEEE Transactions on Instrumentation and Measurement, 2021, (Volume:70) 10.1109/TIM.2021.3109732.

[34] J. Liu, W. Yan, Y. Dong, "Dynamic Hand Gesture Recognition Based on Signals from Specialized Data Glove and Deep Learning Algorithms", IEEE Transactions on Instrumentation and Measurement,2021, (Volume:70).

[35] Z. Wang, T. Zhao, J. Ma, H. Chen, K. Liu, H. Shao, Q. Wang, J. Ren, "Hear Sign Language: A Real-time End-to-End Sign Language Recognition System", IEEE Transactions on Mobile Computing, 2020.

[36] X. Deng and S. Li, "An Improved SSD Object Detection Algorithm Based on Attention Mechanism and Feature Fusion", Journal of Physics: Conference Series 2450 012088, 2023.

[37] L. Wei, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. C. Berg, "SSD: Single Shot MultiBox Detector", European Conference on Computer Vision, 2016, pp 21–37.

[38] Website- Edge Impulse, (2022), Available:https://docs.edgeimpulse.com/docs/edge-impulse-studio/learning-blocks/object-detection/mobilenetv2-ssd-fpn