The Isolated Sign Language Recognition by Accumulative Video Motion

Mohana Sundari M.
M.E, Computer science and engineering,
Velalar college of engineering and Technology, Erode–638012

Arun Pandian M R
Computer science and engineering,
Velalar college of engineering and Technology, Erode–638012

Janani Sri R
Computer science and engineering,
Velalar college of engineering and Technology, Erode–638012

Karan S N
Computer science and engineering,
Velalar college of engineering and Technology, Erode–638012

Abstract -This paper represent Deep Learning based Sign language recognition that helps for primary communication medium between normal persons and hearing impairments. This language depends mainly on hand articulations accompanied by no manual gestures. Recently, there has been a growing interest in sign language recognition. In this paper, we propose a trainable deep learning network for isolated sign language recognition, by using Yolo v3 algorithm which can effectively isolated the sign language from the video frame. Sign Language Recognition system takes an input expression from the hearing impaired person and gives output to the normal persons in the form of text or voice.
Keyword: **Sign Language Recognition, Deep Learning, Yolo v3**

I.INTRODUCTION

Sign language recognition has gained a lot of attention in recent years for automatic explanations by computer or a robot. There are many application scenarios of this research, which can help deaf-dumb people communicate with others in public areas such as hospitals, banks, and train stations. The development of this research will greatly reduce the inconvenience of deaf-dumb people's lives. There are millions of deaf-dumb people in the world communicating by sign language, the communication between the normal people and hearing impaired people is complicated for some times so we designing a Sign Language Recognition system tis very meaningful and valuable for normal people to understand them. In this paper, we investigate a real-time American Sign Language recognition system and we propose a technique to extract key postures for handling the variations in the sign samples performed by different signers. Sign language interpretation methods aim at automatically translating sign languages. Such a process involves mainly two tasks, namely, word-level sign language recognition and sentence-level sign language recognition. In this paper, we target at word-level recognition task for American Sign Language (ASL) considering that it is widely adopted by deaf communities. To recognize sign language using accumulative video motion and deep learning, large datasets of sign language videos can be annotated with labels indicating which signs or gestures are being performed. Deep learning models can then be trained on these datasets to recognize the individual signs or gestures. Recent research has focused on using a combination of CNNs and RNNs, or a 3D-CNN, to analyses both the static and temporal features of sign language video data. These deep learning models can recognize both non-abstract and abstract signs, which can be particularly challenging since abstract signs may not have a clear corresponding word or meaning in spoken language. The isolated sign language recognition using accumulative video motion and deep learning has the potential to improve communication and accessibility for people who use sign language. Further research and development are needed to improve the accuracy and robustness of these models for real-world applications. The isolated sign language recognition by accumulative video motion using deep learning is a promising approach for improving the accuracy and performance of sign language recognition systems.

Figure 1: ASL sign of “Alphabet”

II. LITERATURE SURVEY

A analysis of the literature for the proposed framework reveals that many attempts have been made to tackle sign recognition in videos and images using various methods and algorithms.

[1]. Tse-Yu Pan, Li-Yun Lo, Chung-Wei Yeh, Jhe-Wei Li, Hou-Tim Liu, Min-Chun Hu published the paper real time Sign Language Recognition. Three kinds of features are combined to describe the contours and the salient points of hand gestures. Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Support Vector Machine (SVM) are integrated to construct a novel hierarchical classification scheme. The average recognition accuracy for the CSL dataset is 90%. As for the ASL dataset, we also achieve the recognition accuracy of 85%, which is better than the stacked denoising sparse autoencoders method.

[2].Salem Ameen, Sunil Vadera published A convolutional neural network to classify American Sign Language fingerspelling from depth and color images this paper explores the applicability of deep learning for interpreting sign language and develops a convolutional neural network aimed at classifying fingerspelling images using both image intensity and depth data. The developed convolutional network is evaluated by applying it to the problem of fingerspelling recognition for American Sign Language. The evaluation shows that the developed convolutional network performs better than previous studies and has precision of 82% and recall of 80%. Analysis of the confusion matrix from the evaluation reveals the underlying difficulties of classifying some particular signs, which are discussed in the paper.

[3].Rise Shaw Mary Thuman published 3D Convolutional Neural Network was used to recognize 25 gesture from an Arabic Sign.The value of ASL academic papers for students, both personally and academically, became apparent almost immediately. We saw growth in the development of students’ critical thinking skills and language skills as a result of the process of learning to produce academic ASL papers. We also saw fundamental changes in students’ self-assessment skills and in their perception of ASL. The process of constructing ASL academic papers helped students recognize and address some of their unwitting audism and paternalism. This helped them uncover their unconscious assumptions about ASL.

[4] Ahmad Firooz Shokoori; Masihullah Shinwari published a Sign Language Recognition and Translation into Pashto Language Alphabets**.** Sign language is mandatory and mostly used by huge number of individuals (particularly disabled) in society and it is another way of communication; The form of the hand, the contour of the movement, and the posture of the hand, face, and body parts are all distinct in each sign language; as a result, visual sign language identification is a challenging area of computer vision research; various academics have suggested several models in recent years that have been greatly improved using deep learning approaches. This study looked at the vision-based sign language recognition models suggested utilizing deep learning approaches in the last five years. Although the suggested model’s general trend reveals that sign language recognition accuracy has substantially increased, specific problems still have to be overcome. There are various and multiple languages in the world, people from different languages, have their alphabets and sign to communicate with each other.

III.PROPOSED METHODOLOGY

2: System Architecture

3.1HUMAN POSE ESTIMATION

Human Pose Estimation is a way of identifying and classifying the joints in the human body. Essentially it is a way to capture a set of coordinates for each location key point which is known as a key point that can describe a pose of a person. By capturing various movement and postures from environment different signs can be identified. Human pose estimation has been utilized in a wide range of applications, including human-computer interaction, action recognition, motion capture, movement analysis, augmented reality, sports and fitness, and robotics.

3.2 FEATURE EXTRACTION

Predefined features such as form, contour, geometrical feature (position, angle, distance, etc., color feature, histogram, and others are extracted from the preprocessed images and used later for sign classification or recognition. Feature extraction is a step in the dimensionality reduction process that divides and organizes a large collection of raw data reduced to smaller, easier-to-manage classes. As a result, processing would be simpler. The fact that these massive data sets have a large number of variables is the most important feature. To process these variables, a large amount of computational power is needed. As a result, function extraction aids in the extraction of the best feature from large data sets by selecting and combining variables into functions reducing the size of the data these features are simple to use while still accurately and uniquely describing the actual data collection.

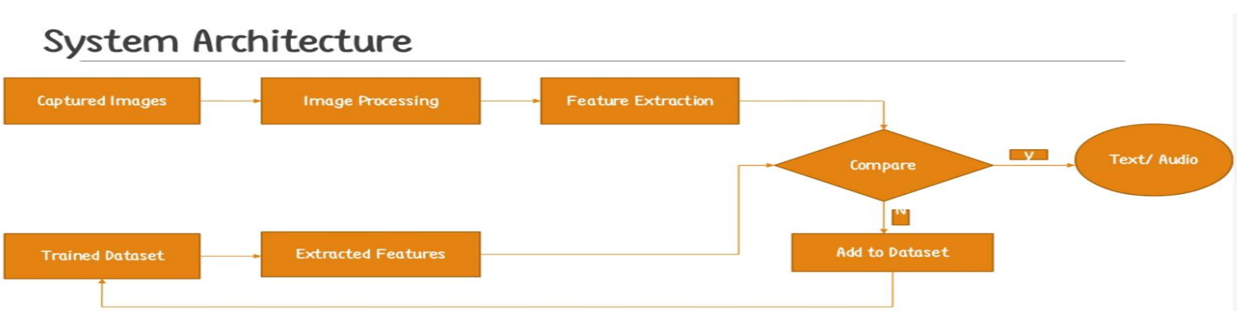
3.3. TEMPORAL MODELLING AND CLASSIFICATION

3.3.1 TEMPORAL MODELLING

Temporal Modelling is a way of modelling software systems and components by putting events first. The usual way of modelling software is to find structures, things and relations. We try to find the relevant aspects of a domain and put all properties into an object-oriented model. Trying to create a second model for a related business process, having the structural model already in place, might result in a process representation that is tightly coupled with the assumptions built up from the structural model and too far away from reality.

3.3.2CLASSIFICATION

Using the temporal model to classify the sign language gesture into its corresponding meaning. Here we using a various classification algorithms, such SVM or decision trees.Classification is a deep learning method where the model tries to predict the correct label of a given input data. In classification, the model is fully trained using the training data, and then it is evaluated on test data before being used to perform prediction on new unseen data.

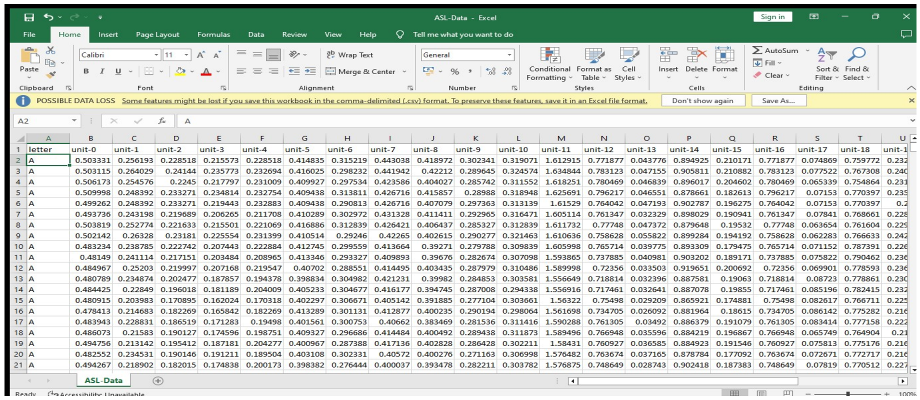


3.4FINAL STAGE OF PROCESSING

In the final stage of processing, the isolated sign language recognition by accumulative video motion is usually the output in the form of text or audio from the interpretation of sign language gesture. Once the model has been developed and the classification algorithm applied, the system produces an output that corresponds to the meaning of the sign language gesture. Understanding human behavior and identifying various postures and body movements, as well as translating them into text.

IV.METHODOLOGY USED

4.1 TRAINED DATASET



letter	unit-0	unit-1	unit-2	unit-3	unit-4	unit-5	unit-6	unit-7	unit-8	unit-9	unit-10	unit-11	unit-12	unit-13	unit-14	unit-15	unit-16	unit-17	unit-18	unit-1
A	0.503331	0.256193	0.228518	0.215573	0.228518	0.414835	0.315219	0.443038	0.418972	0.302341	0.319071	1.612915	0.771877	0.043776	0.894925	0.210171	0.771877	0.074869	0.759772	0.232
A	0.503115	0.264029	0.24144	0.235773	0.232694	0.416025	0.298232	0.441942	0.42212	0.289645	0.324574	1.634844	0.783123	0.047155	0.905811	0.210882	0.783123	0.077522	0.767308	0.240
A	0.506173	0.254576	0.2245	0.217797	0.231009	0.409927	0.297534	0.423586	0.404027	0.285742	0.311552	1.618251	0.780469	0.046839	0.896017	0.204602	0.780469	0.065339	0.754864	0.231
A	0.509998	0.248392	0.233271	0.234814	0.232754	0.409438	0.313811	0.426716	0.415857	0.28988	0.318948	1.625691	0.796217	0.046551	0.878661	0.182613	0.796217	0.07153	0.770397	0.235
A	0.499262	0.248392	0.233271	0.219443	0.232883	0.409438	0.290813	0.426716	0.407079	0.297363	0.313139	1.61529	0.764042	0.047193	0.902787	0.196275	0.764042	0.07153	0.770397	0.2
A	0.493736	0.243198	0.219689	0.206265	0.211708	0.410289	0.302972	0.431328	0.411411	0.292965	0.316471	1.605114	0.761347	0.032329	0.898029	0.190941	0.761347	0.07841	0.768661	0.228
A	0.503819	0.252774	0.221633	0.215501	0.221069	0.416886	0.312839	0.426421	0.406437	0.285327	0.312839	1.611732	0.77748	0.047372	0.879648	0.19532	0.77748	0.063654	0.761604	0.229
A	0.502142	0.26328	0.23181	0.225554	0.231399	0.410514	0.29246	0.42265	0.402615	0.290277	0.321463	1.610636	0.758628	0.055822	0.899284	0.194192	0.758628	0.062283	0.766633	0.242
A	0.483234	0.238785	0.222742	0.207443	0.222884	0.412745	0.299559	0.413664	0.39271	0.279788	0.309839	1.605998	0.765714	0.039775	0.893309	0.179475	0.765714	0.071152	0.787391	0.226
A	0.48149	0.241114	0.217151	0.203484	0.208965	0.413346	0.293327	0.409893	0.39676	0.282674	0.307098	1.593865	0.737885	0.040981	0.903202	0.189171	0.737885	0.075822	0.790462	0.236
A	0.484967	0.25203	0.219997	0.207168	0.219547	0.40702	0.288551	0.414495	0.403435	0.287979	0.310486	1.589998	0.72356	0.033503	0.919651	0.200692	0.72356	0.069901	0.778593	0.236
A	0.480789	0.234874	0.202477	0.187857	0.194378	0.398834	0.304982	0.421231	0.39982	0.284853	0.303581	1.556649	0.718814	0.032396	0.887581	0.19063	0.718814	0.08723	0.778861	0.230
A	0.484425	0.22849	0.196018	0.181189	0.204009	0.405233	0.304677	0.416177	0.394745	0.287008	0.294338	1.556916	0.717461	0.032641	0.887078	0.19855	0.717461	0.085196	0.782415	0.232
A	0.480915	0.203983	0.170895	0.162024	0.170318	0.402297	0.306671	0.405142	0.391885	0.277104	0.303661	1.56322	0.75498	0.029209	0.865921	0.174881	0.75498	0.082617	0.766711	0.225
A	0.478413	0.214683	0.182269	0.165842	0.182269	0.413289	0.301131	0.412877	0.400235	0.290194	0.298064	1.561698	0.734705	0.026092	0.881964	0.18615	0.734705	0.086142	0.775282	0.216
A	0.483943	0.228831	0.186519	0.171283	0.19498	0.401561	0.300753	0.40662	0.383469	0.281536	0.311416	1.590288	0.761305	0.03492	0.886379	0.191079	0.761305	0.083414	0.777158	0.222
A	0.486073	0.21583	0.190127	0.174596	0.198751	0.409327	0.296686	0.414484	0.400492	0.289438	0.311873	1.589496	0.766948	0.035596	0.884219	0.196867	0.766948	0.065749	0.764904	0.21
A	0.494756	0.213142	0.195412	0.187181	0.204277	0.400967	0.287388	0.417136	0.402828	0.286428	0.302211	1.58431	0.760927	0.036585	0.884923	0.191546	0.760927	0.075813	0.775176	0.216
A	0.482552	0.234531	0.190146	0.191211	0.189504	0.403108	0.302331	0.40572	0.400276	0.271163	0.306998	1.576482	0.763674	0.037165	0.878784	0.177092	0.763674	0.072671	0.772717	0.216
A	0.494267	0.218902	0.182015	0.174838	0.200173	0.398382	0.276444	0.400037	0.393478	0.282211	0.303782	1.576875	0.748649	0.028743	0.902418	0.187383	0.748649	0.07819	0.770512	0.227

Figure 3: Trained Dataset

4.2 DATASET DESCRIPTION

We analyse 1026 images of sign images which is ASL images of Alphabet which have a spread of 26 class labels assigned to them. Each class label is a set of sign images of English Alphabet. All the images are resigned to 640 x 480 pixel and we perform both the model optimization and prediction on these downscaled images

4.3 EXTRACTED FEATURE

The process of choosing and modifying the input characteristics that will be used to generate predictions is known as feature extraction, and it is a crucial stage in the development of a machine learning model. Finding the characteristics that will have the most influence on the target variable and are the most pertinent and informative is the aim of feature extraction. This is the procedure for selecting the model's most relevant attributes. This can be accomplished using methods like correlation analysis, principal component analysis, or feature significance estimations using decision-tree-based algorithm.

4.4 EXPERIMENTAL RESULT

The extracted feature in a dataset is compared with the feature extraction of captured video frame gestures. As a result, comparing both the features are same then it shows the result in the form of text or audio. On the other hand it cannot be same to the dataset then it automatically move and create the new dataset.

V. CONCLUSION

In conclusion, isolated sign language recognition by accumulative video motion using deep learning is a promising approach for improving the accuracy and performance of sign language recognition systems. This approach involves using deep learning algorithms, to analyse and interpret the motion information from video frames. The accumulative video motion approach has shown good results in recognizing isolated sign language gestures, which can be useful for applications such as assistive technology and communication aids for the deaf and hard of hearing. By improving the accuracy and performance of sign language recognition systems, this approach can also help to bridge the communication gap between hearing and deaf communities. There are still challenges that need to be addressed, such as the high variability and complexity of sign language gestures and the need for large datasets for training deep learning models. Continued research in this area is important to further improve the accuracy and performance of sign language recognition systems and to make them more accessible and useful for people who rely on sign language to communicate.

VI. FUTURE WORK

There are several potential areas for future enhancement in isolated sign language recognition by accumulative video motion using deep learning.

- Improving the accuracy of the recognition system
- Addressing the variability and complexity of sign language gestures
- Addressing the variability and complexity of sign language gestures
- Enhancing real-time performance

REFERENCES

[1] Muhammad Reza Jafari and Dr. Vinod Kumar, "Persian Sign Gesture Translation to English Spoken Language on Smartphone", *INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH TECHNOLOGY (IJERT),* vol. 09, no. 06, June 2020.

[2] M. Lee and J. Bae, "Deep Learning Based Real-Time Recognition of Dynamic Finger Gestures Using a Data Glove", *IEEE Access*, vol. 8, pp. 219923-219933, 2020

[3] Nagarajan C., Neelakrishnan G., Akila P., Fathima U., Sneha S. "Performance Analysis and Implementation of 89C51 Controller Based Solar Tracking System with Boost Converter" Journal of VLSI Design Tools & Technology. 2022; 12(2): 34–41p.

[4] C. Nagarajan, G.Neelakrishnan, R. Janani, S.Maithili, G. Ramya "Investigation on Fault Analysis for Power Transformers Using Adaptive Differential Relay" Asian Journal of Electrical Science, Vol.11 No.1, pp: 1-8, 2022.

[5] G.Neelakrishnan, K.Anandhakumar, A.Prathap, S.Prakash "Performance Estimation of cascaded h-bridge MLI for HEV using SVPWM" Suraj Punj Journal for Multidisciplinary Research, 2021, Volume 11, Issue 4, pp:750-756

[6] G.Neelakrishnan, S.N.Pruthika, P.T.Shalini, S.Soniya, "Perfromance Investigation of T-Source Inverter fed with Solar Cell" Suraj Punj Journal for Multidisciplinary Research, 2021, Volume 11, Issue 4, pp:744-749

[7] S. Kido, Y. Hirano and N. Hashimoto, "Detection and classification of lung abnormalities by use of convolutional neural network (CNN) and regions with CNN features (R-CNN)", *2018 International Workshop on Advanced Image Technology (IWAIT)*, pp. 1-4, 2018.

[8] "Sign Language Recognition for Computer Vision Enthusiasts", *Data Science Blogathon*, [online] Available: https://www.analyticsvidhya.com/blog/2021/06/sign-language-recognition-for-computer-vision-enthusiasts/.

[9] S Shivashankara and Dr. Srinath, "American Sign Language Recognition System: An Optimal Approach", *International Journal of Image Graphics and Signal Processing.* vol. 10, 2018.

[10] Murat KıllıoğluTaskiran and Nihan Mehmet Kahraman, "A Real-Time System for Recognition of American Sign Language by Using Deep Learning", 2018.

[11] Y. Saleh and G. F. Issa, "Arabic Sign Language Recognition through Deep Neural Networks Fine-Tuning", *Int. J. Onl. Eng.*, vol. 16, no. 05, pp. 71-83, May 2020

[12] Y. Yan, Z. Li, Q. Tao, C. Liu and R. Zhang, "Research on Dynamic Sign Language Algorithm Based on Sign Language Trajectory and Key Frame Extraction", *2019 IEEE 2nd International Conference on Electronics Technology (ICET)*, pp. 509-514, 2019

[13] O. M. Sincan and H. Y. Keles, "AUTSL: A large scale multi-modal turkish sign language dataset and baseline methods", *IEEE Access*, vol. 8, pp. 181 340-181 355, 2020.

[14] S. Jiang, B. Sun, L. Wang, Y. Bai, K. Li and Y. Fu, "Skeleton aware Multi-modal sign language Recognition", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3413-3423, 2021.

[15] C.Nagarajan and M.Madheswaran - 'Experimental verification and stability state space analysis of CLL-T Series Parallel Resonant Converter' - Journal of ELECTRICAL ENGINEERING, Vol.63 (6), pp.365-372, Dec.2012.

[16] C.Nagarajan and M.Madheswaran - 'Performance Analysis of LCL-T Resonant Converter with Fuzzy/PID Using State Space Analysis'- Springer, Electrical Engineering, Vol.93 (3), pp.167-178, September 2011.

[17] C.Nagarajan and M.Madheswaran - 'Stability Analysis of Series Parallel Resonant Converter with Fuzzy Logic Controller Using State Space Techniques'- Taylor & Francis, Electric Power Components and Systems, Vol.39 (8), pp.780-793, May 2011.

[18] Nagarajan and M.Madheswaran - 'Experimental Study and steady state stability analysis of CLL-T Series Parallel Resonant Converter with Fuzzy controller using State Space Analysis'- Iranian Journal of Electrical & Electronic Engineering, Vol.8 (3), pp.259-267, September 2012