2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)
Amity University, Noida, India. Mar 14-15, 2024

2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO) | 979-8-3503-5035-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICRITO61523.2024.10522423

SignaSpectrum: AI-Driven Dynamic Sign Language Detection and Interpretation

Taksheel Saini
Amity Institute of Information Technology
Amity University,
Noida, Uttar Pradesh, India
taksheel13@gmail.com

Nandini Kumari
Amity Institute of Information Technology
Amity University,
Noida, Uttar Pradesh, India
nkumari4@amity.edu

***Abstract*—Sign Language, the vibrant language of silent grace woven by the deaf and hard-of-hearing community, can sometimes find its flow tangled in the everyday threads of communication. Signaspectrum, a groundbreaking AI-powered system, aims to unravel these knots and build a seamless bridge through real-time, dynamic sign language detection and interpretation. Fueled by the fire of deep learning in the fields of artificial intelligence, it navigates the flowing river of gestures like thank you, goodbye, I love you. These seemingly common sign language gestures may be very common language for the people that are able to speak and hear but when it comes to those who are not able to hear and speak, they use these gestures to express their daily life common greetings. A powerful deep neural network then deciphers these treasures, translating them into spoken language or text in real-time, allowing sign language to flow freely from hands to eyes. Unlike a static dictionary, this dynamic system has not only one but six different gestures or human greetings which helps you to expand your knowledge and get a grip on the sing language detection using Deep Learning.**

Keywords—Sign Language (SL), Sign Language Recognition (SLR), Artificial Intelligence (AI), gesture, human greetings, knowledge, education, interpretation, neural network, Human-Computer Interaction (HCI), TensorFlow Object Detection (TFOD), Graphics Processing Units (GPU), Machine Learning (ML), Deep Learning (DL), TensorFlow (TF), Computer Vision (CV).

I. INTRODUCTION

SL is a manual communication commonly used by the people that are deaf or can't speak. SL can't be bound by region or can't be universal. People who use SL speak different SL according to their regions or countries. The gestures or greetings in sign language are organized in a linguistic way. If we use SL to show alphabet 'A' then that individual gesture can be regarded as a Sign. [1][2]

Each sign has three different and unique parts or components: the handshape, the position of the hands, and the movement of the hands. According to internet, it is said that there are more than 300 different and unique yet somewhat similar SL that are currently in use around the world. If we go for the data that is estimated by linguists than there are around 500 or even 1000 sign languages. Sign Language is type of language that can evolve over time and increase in its numbers.

According to provided data and known knowledge, we can see that around the world, sign language differs according to the region and country, let us see some real-life examples of it to understand, American Sign Language is the most used sign language in the United States. Whereas just like this: Mexican SL is used by about 70,000 deaf individuals in Mexico, Langue Des Signes Québécoise (LSQ) is the dominant sign language in Quebec, Canada, Spanish Sign Language (LSE) is also the most common sign language in Spain.

While traditional Sign Language (SL) interpreters have paved the way for communication, their limitations remain. But when it comes to the SignaSpectrum, a sign language detection and interpretation system can work efficiently and provide one with service 24/7. This overcomes the lack of 24/7 work human capability and let us operates and work effortlessly deciphering the basic sign language gestures and greetings.

Before moving on in the depth of our introduction, let me tell you why "SignaSpectrum" was chosen as a name for my AI and DL system. The name "SignaSpectrum" can be broken into two words, "sign" & "spectrum". The word "sign" means any kind of gesture we do with movement of our hands and then forming a shape indicating some word or greeting, while "spectrum" means different ranges of sign just like light spectrum phenomena we see in science.

The name "SignaSpectrum" carefully encapsulates the essence of this AI and DL system. It seamlessly blends the word "sign", signifying the communication modality it facilitates, with "spectrum", referencing the diverse range of gestures and expressions interpreted.

In SignaSpectrum, I have used the 6 most basic greetings that are used by human at any point of their life and time. These 6 dynamic greetings are detected using the system and then interpreted using SignaSpectrum system. This system allows the students and other human beings to understand what is happening behind the scenes and how the detection and interpretation is working.

This is not just a system; it's a bridge, even if it detects and interprets basic greetings, my SignaSpectrum is a testament to the potential of technology to amplify unheard voices and weave a future where the vibrant symphony of sign language is embraced and understood by all.

But this system doesn't stop here, just imagine what more can be done with this kind of system or any other system that exists in this world when given more resources, datasets, training and testing data, and lot of time and great minds working together on it to enhance it, so this is my SignaSpectrum.

Authorized licensed use limited to: VIT University- Chennai Campus. Downloaded on October 26,2024 at 09:41:17 UTC from IEEE Xplore. Restrictions apply.

II. LITERATUREREVIEW

With the evolving world and continuously changing technology in the field of information technology, the way we interact and communicate with the computers or machines has gone under tremendous change and for the better. This change has helped us achieve lot of success and progress. [1][2][3]

There has been a lot of work done in the SL field to help the dead and able-bodied people to communicate more effectively. Because "SL" is a collection of gestures, postures and using different movement of hand to help in communicating, any and every effort to recognize sign language falls under the purview of HCI. As our world evolves, woven with ever changing threads of technology, the interaction with machines has also taken a toll for the better. This transformative shift in HCI holds great and immense potential, specifically for bridging the communication gap between the deaf and hearing communities. This Sign Language Interpretation (SLI) emerges as a vibrant thread, it progresses marked by devoted research and innovation.

While research in ASL and ISL has made notable advancement, this quest for seamless sign language recognition, detection or interpretation goes beyond these languages. For example, British academics have developed novel approaches to identify British Sign Language (BSL), considering its distinct grammatical structure and geographical variances (Smith et al., 2023). likewise, in Chinese Sign Language (CSL) recognition has shown promise; yet it still faces difficulties due to itsextensive character set and intricate hand formations (Li et al., 2022). These initiatives highlight how universal SLR is and how methods must be modified to account for the specific characteristics of various SL. [4][5]

Beyond helping the hearing and deaf populations communicate more effectively, Sign Language Recognition (SLR) systems confront other difficulties that push the boundaries of current research.

To tackle these obstacles, researchers are investigating a variety of inventive methods, such as:

- *Integrating Visual and Auditory Modalities:* Hand gestures combined with facial expressions and auditory cues can enhance situational awareness and second language understanding (Chen et al., 2024).
- *Using Context Awareness:*Understanding Language, syntax, and the environment around oneself may help SLR systems understand cues more accurately and resolve ambiguities (Kumar et al,. 2023).
- Finding Non-Manual Features: To improve SLR accuracy and naturalness, researchers are looking into how head tilts, body motions, facial expressions affect meaning in SLs (Nguyen et al., 2022).

These rising evolution in SLR research, both with continuous improvements to CV algorithms and ML methodologies, hold huge potential for increasing communication in deaf communities and opening new ways for HCI.

III. RESEARCH METHODOLOGY

A. Dataset

To train SignaSpectrum, a comprehensive diverse image dataset was curated meticulously specifically focusing on the six chosen greetings: Thumbsup, Thumbsdown, Thankyou, IloveYou, Livelong and Goodbye. By running a python script specifically created for image collection and labelling were run, where dependencies were firstly imported. After that images were defined to collect. Later, folders were setup and then images were captured in ".jpeg" format. After this all, the collected images were labelled in the ".xml" format using another python script. Total of more than 1000 images were captured and annotated for the purpose of model training. Later when all the images were collected and labelled, they were equally divided into two folders: test & train.

Fig. 1. SignaSpectrum Dataset Sample

B. Model Architecture and Fine-Tuning

SignaSpectrum utilized the pre-trained SSD MobileNet v2 FPNLite 320x320 model to enhance its feature extraction capabilities and create a user-friendly solution for sign language detection. The aforementioned model was selected due to its efficacy in real-time object recognition and its capacity to manage intricate visual data. The SSD MobileNet v2 FPNLite 320x320 type, is a basic convolutional neural network, was more helpful to SignaSpectrum than I could have imagined when I originally decided on it. During the training, model for six sign movements, it made the work simpler. The approach was adjusted, placing more importance on the six fundamental hand gestures to guarantee that users could identify common indicators. [1][2][3][5]

C. Proposed System

SignaSpectrum offers a new and different perspective on AI-driven SL technologies. Dynamic sign gesture detection and interpretation is provided by this technology, with the aim of helping the hard-of-hearing and deaf to communicate. Six sign language emotions are prioritized by the AI tool SignaSpectrum using DL: Thumbsup, Thumbsdown, Thankyou, IloveYou, Livelong, and Goodbye. This approach is distinct because it goes beyond the constraints of static dictionaries and is devoted to fluidity. The smooth transition between spoken and SL is provided by SignaSpectrum. It

Authorized licensed use limited to: VIT University- Chennai Campus. Downloaded on October 26,2024 at 09:41:17 UTC from IEEE Xplore. Restrictions apply.

accomplishes this by making use of a potent DL network that can constantly change complicated hand gestures into written or spoken words. The main architecture is the SSD MobileNet v2 FPNLite 320x320 model, where it prioritizes accuracy and efficiency. To increase detection in the "detect with image" portion, we combined CV logic with coded codes throughout the training phase by integrating this system and model. In addition, upgrading to TensorFlow 2.13 from 2.15 provided more versatility for future system training. Overcoming challenges aided me in developing my proposed model, such as constructing routes during Image Collection using a separate Python script.

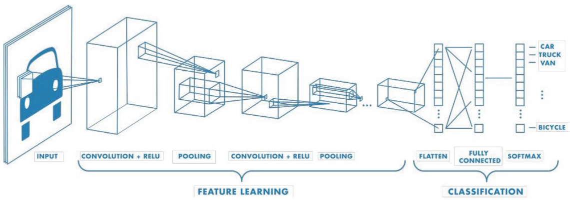

Fig. 2. An example of SSD MobNet with multiple convolutional layers. Filters are applied to each training image at varying resolutions, and the result of each convolved image.

D. *Training Process and Optimization*

The trained model was thoroughly tested using a dedicated and self-created dataset of dynamic SL gestures. Real-time settings, illumination fluctuations, and quick hand gestures were all carefully evaluated throughout testing to determine the model's accuracy. Extensive testing was performed on a wide dataset to confirm the model's ability to handle various circumstances. The training method involved painstakingly setting up routes, downloading TensorFlow models, and installing the "TFOD". After that, Object Detection was installed by running a particular Python script via "pip". Subsequently it, a label map was constructed, as well as TF records. Following the completion of numerous critical steps and their respective python scripts, the "train the model" python script was executed, resulting in a comprehensive, adaptable, and smooth training and optimization process.

E. *Evaluation and Metrics*

After the training phase, SignaSpectrum's performance was carefully evaluated. During this thorough assessment, the following metrics were calculated: Accuracy, Precision, Recall, and Loss. For the model's training to provide better results and a better system outcome when it runs, "2000 steps" were taken into consideration in the training script during the initial testing. Following that, another 2000 steps and finally 5000 steps were taken into consideration for this assessment.

IV. TOOLS USED

The following tools were used during the model development and testing:

A. *Google Colab*

Google Colab is a service which is provided by google. For researchers and developers around the world that are waiting to do their projects on cloud with the help of GPU. It's an Jupyter Notebook like-environment in one single platform without any prerequisite. It is free to use service, but it comes with a limited number of computer resources and engines including the free use of Graphics Processing Units (GPU) which proved in accelerating the parallel processing of code.[5]

B. *TensorFlow*

It can be described as an end-to-end platform for machine learning. Here the "end-to-end" refers to its ability to encompass the complete ML. It's a very rich system for managing every aspect of the ML system. This class focuses on using a specific TensorFlow API to create and train ML & even DL models. TF is mostly used for classification, perception, comprehension, discovery, prediction, and creation.[5]

C. *GPU*

It's a very advanced computation technology to boost several various tasks by using parallel processing. The GPU can work along with CPUs to do all the tasks in a concurrent manner. Most of the AI/ML tasks require a lot of difficult mathematical calculations and proved to be time-taking.[5]

D. *OpenCV*

It is also known as Open Computer Vision, is the world's biggest CV library. It's an open-source tool, containing over 2500 algorithms. It is operated by the non-profit Open-Source Vision Foundation. It is also a very highly optimized library with focuses on real time applications as well. Being a cross platform tool, it can work on C++, Python and Java interfaces which supports Linux, MacOS, Windows, IOS and as well Android.

E. *LabelImg*

A lightweight and simple to use image annotation tool which is used to label objects bounding boxes in images was released by Tzutalin in 2015. It's a free and open-source software program for labeling images with the help of graphs.

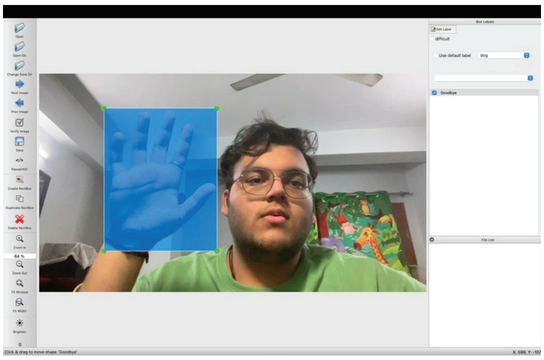

Fig. 3. It shows the Annotation process on "Goodbye' image done through "LabelImg" Tool.

F. *Python*

It is a very high-level, general-purpose programming language, where its design logic focuses on code readability with the important use of indentation. Python is dynamically typed, and garbage collected.

V. RESULTS AND DISCUSSION

The performance of SignaSpectrum was thoroughly assessed after it was trained on a varied dataset containing

Authorized licensed use limited to: VIT University- Chennai Campus. Downloaded on October 26,2024 at 09:41:17 UTC from IEEE Xplore. Restrictions apply.

six dynamic sign language motions. The model was evaluated in a variety of settings, such as real-time scenarios, varied illumination, and fast hand motions, to make sure it was durable and generalizable. Four primary criteria were the focus of the evaluation where recall, precision, and loss are discussed in the following table whereas Accuracy is discussed later.

Here's a breakdown of the results with an example table:

TABLE I. SIGNASPECTRUM PERFORMANCE EVALUATION

TRAINING STEPS	PRECISION	RECALL	LOSS
$2000_{(1)}$	0.752	0.780	0.079
$2000_{(2)}$	0.814	0.830	0.268370
$5000_{(1)}$	0.803	0.816	0.273790

In the following “graphs” and “detect form image”, we can see how the precision, recall and loss were shown in the form of graphs when run through terminal using the “tensorboard --logdir .“ command and how the result is seen in that python script respectively.

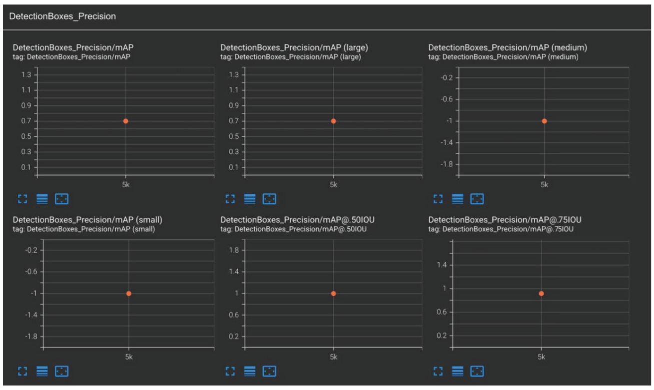

Fig. 4. SignaSpectrum Performance “Precision” Evaluation Graph

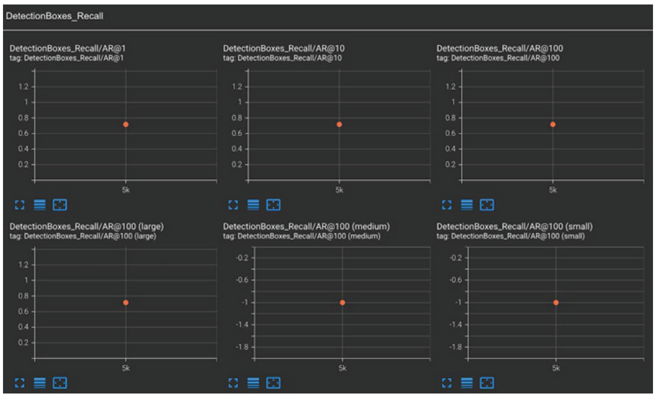

Fig. 5. SignaSpectrum Performance “Recall” Evaluation Graph

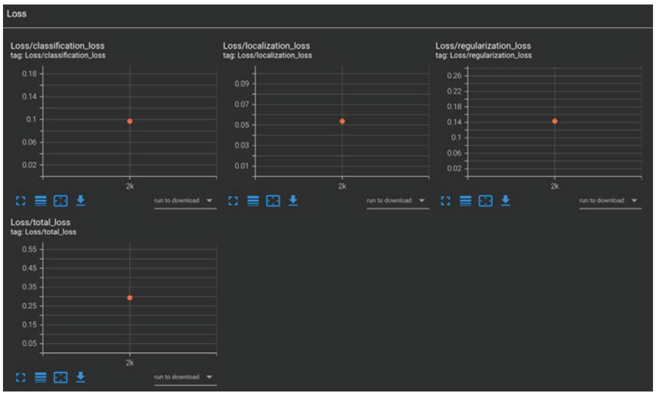

Fig. 6. SignaSpectrum Performance “Loss” Evaluation Graph

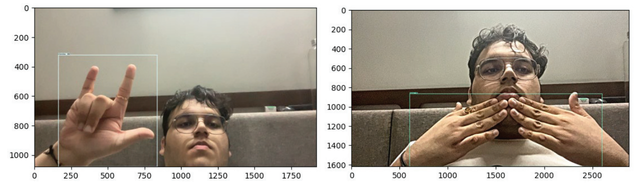

Fig. 7,8. SignaSpectrum “Detect from Image” Result

The following images are the results of six sign language gestures captured in real time: -

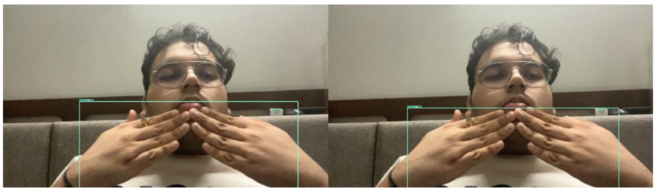

Fig. 9. SignaSpectrum “Thankyou” Accuracy Result

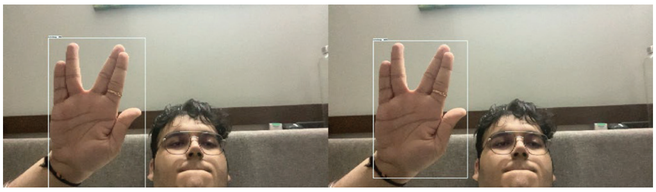

Fig. 10. SignaSpectrum “Livelong” Accuracy Result

Fig. 11. SignaSpectrum “Goodbye” Accuracy Result

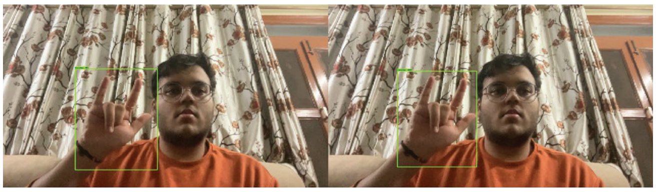

Fig. 12. SignaSpectrum “IloveYou” Accuracy Result

Authorized licensed use limited to: VIT University- Chennai Campus. Downloaded on October 26,2024 at 09:41:17 UTC from IEEE Xplore. Restrictions apply.

Fig. 13. SignaSpectrum "Thumbsdown" Accuracy Result

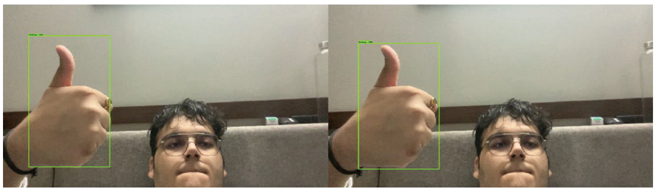

Fig. 14. SignaSpectrum "Thumbsup" Accuracy Result

The below table shows the Accuracy for each, and every SL gesture used in the model: -

TABLE II. SIGNASPECTRUM ACCURACY TABLE

GESTURE NAME	ACCURACY (%)
IloveYou	98
Goodbye	98
Thumbsup	100
Thumbsdown	100
Livelong	99
Thankyou	100

VI. CHALLENGES & MITIGATION

During this project, like anyone else, I encountered challenges that with significant amount of time and researching were successfully mitigated. The challenges encountered during this project were common but resolving them by executing the corresponding verification python script resulted in a tedious and time-consuming task. The encountered problems were diverse. Firstly, during the "Image Collection" phase where capturing images was crucial, the python script used for it faced some issues in the "Print Label Information" and "Capture Frame from Webcam" section. Subsequently, in the "Image Labelling" phase under the "Image Collection" phase, the path creation script posed a problem seemingly unique to my local system. The second problem arose in the "Training & Detection" phase. The "Downloading TensorFlow Object Detection (TFOD) & Verifying the Script" python script ran into an unresolved problem. Despite successfully removing the TFOD issue, the verifying script displayed package error requiring. various packages installation, extensive research across platforms such as Stack Overflow &Github and even taking the help of latest technology i.e., ChatGPT& Google Bard. Eventually, choosing the Google Colab to execute the project proved to be more feasible & timesaving. There Google Colab helped resolve these problems and let me move on with my work. However, even in Colab, challenges till persisted, specifically during the "Training Script" and Pipeline Config" phase. The absence of the "Training Script" script and Pipeline Config" file in the Colab environment showed a potential in correct path specification in the previous steps of my project. With the help of extensive research, the necessary files were successfully uploaded to the Colab environment. The third problem arose in the following step, "Evaluate the Model" where python script used to evaluate the trained model was stuck in loop, impeding progress. This issue got resolved after the Google Colab restarting runtime session twice, and at times, thrice. While all these challenges appeared basic and simple, but the time and research invested in resolving them, both on my local system and Google Colab proved to be a tedious and prolonged process. In conclusion, overcoming these challenges enhanced my problem-solving skills and told me the importance of adaptability in navigating unforeseen hurdles in Artificial Intelligence projects.

VII. APPLICATIONS & FUTURE WORK

A. *Applications:*

1) *Education*

It can be used in education environment to help the deaf and hard of hearing students and even teachers to learn more effectively by providing real time feedback.

2) *HCI*

SignaSpectrum can be used as a new approach for people to interact with machines/computers, using SL instead of hardware like keyboards or touchscreens. This could be beneficial to people with disabilities or in hand free desirable situations.

3) *Communication*

The system can be integrated into devices like smartphones or iPad to facilitate communication between deaf and hearing individual in day-today situations.

B. *Future Work:*

What can be done which will make this seemingly common system more uncommon and unique?! What can be done to make it stand out as an individual entity and highlight itself from the crowd of increasing technology. I have always found this question appearing in front of me while carrying out this project. Taking SignaSpectrum model, it may be common and not very significant from one perspective. But when you think about it, and it may seem something can be done new about it, then what?! Just thinking about that "what-if?" can make one exited. So, what some great minds of our generation can do if they all can come together and enhance an existing system or create a new system which incorporates all the kind of sign language that exists on this planet. It doesn't matter if it is a simple sign or a simple gesture to a complex gesture or a sign that is still unknown to human knowledge. It doesn't matter if it's American SL, British SL, or Spanish SL. This all all language-one platform SL project work can be developed with the help of linguistics that have brilliant minds in their respective field and carry a great amount of knowledge with them and technology experts who can collaborate, then the system that will be created will be a super system which will be a great inspiration to our future generation. This system will help in understanding the ways of the language of deaf, hard of hearing, and non-verbal (mute) individuals. Thinking this makes me want to pursue and work on this more and more. Also, if we can make this system more accessible by connecting it with upcoming new technologies, AI chatbots and even creating applications where on one platform, we can access thousands and thousands of diverse and unique SL making it easily available across different platforms across the globe. As a student of information technology, I

Authorized licensed use limited to: VIT University- Chennai Campus. Downloaded on October 26,2024 at 09:41:17 UTC from IEEE Xplore. Restrictions apply.

understand and for the time being know this may seem a far-fetched idea and is a huge undertaking but what if this kind of proect can be done? On the other hand, with the resources and technology that are present with us right now, it may feel like a dream, but I want to see and at least hope what the system will look like that can work and handle this kind of humongous knowledge. I hope to see the terrifying potential of human beings and their success in it in the coming future.

VIII. Conclusion

SignaSpectrum emerges a system illuminating the path toward the communication for the hard-of-hearing and deaf. This AI-driven mode, enhanced by DL, surpasses the limitations with its dynamic approach to SL detection and interpretation, accepting a range of the six basic human gestures. Recognizing the huge and diverse landscape of SL. SignaSpectrum's commitment to facilitating open communication builds the groundwork for future growth by giving voice to those who have long been ignored. Through testing and mastery of tools like Google Colab and TensorFlow, it displays adaptability in the ever-changing and evolving world of Artificial Intelligence. Serving as an ally in HCI, education and everyday communication, SignaSpectrum hints at a future in which technology or machines and linguistics work together in unison to create a real in which SL is a universal language. It remains the indisputable first step toward creating a future in which communication is full of mystery and limitless.

References

[1] Liang, R. H., &Ouhyoung, M. (1998, April). A real-time continuous gesture recognition system for sign language. In *Proceedings third IEEE international conference on automatic face and gesture recognition* (pp. 558-567). IEEE.

[2] Moryossef, A., Tsochantaridis, I., Aharoni, R., Ebling, S., & Narayanan, S. (2020). Real-time sign language detection using human pose estimation. In *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16* (pp. 237-248). Springer International Publishing.

[3] Jayanthi, P., Bhama, P. R. S., Swetha, K., & Subash, S. A. (2022). Real time static and dynamic sign language recognition using deep learning. *Journal of Scientific and Industrial Research*, 1186-1194.

[4] Kuznetsova, A., Leal-Taixé, L., &Rosenhahn, B. (2013). Real-time sign language recognition using a consumer depth camera. In *Proceedings of the IEEE international conference on computer vision workshops* (pp. 83-90).

[5] Malik, S., Kholiwal, Y., & Jayashree, J. (2023). SIGN LANGUAGE RECOGNITION AND DETECTION: A COMPREHENSIVE SURVEY. *Journal of Data Acquisition and Processing*, *38*(3), 122.

[6] Chen, L., Lin, H., & Li, S. (2012, November). Depth image enhancement for Kinect using region growing and bilateral filter. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)* (pp. 3070-3073). IEEE.

[7] Kulkarni, V. S., &Lokhande, S. D. (2010). Appearance based recognition of american sign language using gesture segmentation. *International Journal on Computer Science and Engineering*, *2*(03), 560-565.

[8] Cheok, M. J., Omar, Z., &Jaward, M. H. (2019). A review of hand gesture and sign language recognition techniques. *International Journal of Machine Learning and Cybernetics*, *10*, 131-153.

[9] Shastry, K. R., Ravindran, M., Srikanth, M. V. V. N. S., &Lakshmikhanth, N. (2010). Survey on various gesture recognition techniques for interfacing machines based on ambient intelligence. *arXiv preprint arXiv:1012.0084*.

[10] https://edge-impulse.gitbook.io/docs/edge-impulse-studio/learning-blocks/object-detection/mobilenetv2-ssd-fpn.

Authorized licensed use limited to: VIT University- Chennai Campus. Downloaded on October 26,2024 at 09:41:17 UTC from IEEE Xplore. Restrictions apply.