https://doi.org/10.46335/IJIES.2024.9.7.4 **e-ISSN: 2456-3463**
Vol. 9, No. 7, 2024, PP. 16 -20

International Journal of Innovations in Engineering and Science, www.ijies.net

Breaking barriers with SignSpeak: Transforming sign language into text and audio effortlessly

Atharva Wasade, Harshalsingh Solanki, Jayesh Wadaskar, John Koshy

St. Vincent Pallotti College of Engineering & Technology
Gavasi Manapur, Nagpur, Maharashtra, India

jswadaskar2001@gmail.com

***Received on**: 5 May,2024* ***Revised on**: 29 June,2024* ***Published on**: 02 July ,2024*

Abstract— **The objective of this research is to analyze and assess current sign language translation technologies, emphasizing their efficiency, precision, and user-friendliness. It delves into the capacity of these translation tools to enhance interactions within the deaf and hard-of-hearing community. The investigation adopts a thorough methodology to review a range of sign language translation devices, taking into account aspects like instantaneous translation, simplicity of operation, and compatibility with various sign language dialects. The research incorporates hands-on trials and feedback from users to gauge the effectiveness and practicality of these devices. The findings aim to pinpoint both the advantages and shortcomings, thereby offering insights that could steer the advancement of more robust and accessible communication aids for those with auditory and vocal disabilities.**

Keywords— Gesture Recognition , MediaPipe , Tensorflow.

I. INTRODUCTION

Verbal communication, while primary for most, poses Verbal communication is fundamental for most, yet it poses challenges for the mute or deaf. Deafness, from various factors like genetics or aging, hinders auditory communication. Muteness can stem from causes like illness or trauma. Presently, around 466 million globally face hearing impairment, with 34 million being children. By 2050, projections suggest it'll reach 900 million, implying a rise in speech impairment cases. Sign language, with roots dating to 5th century B.C., serves as primary communication. American Sign Language (ASL) is prominent, aiding expression for the deaf and mute. However, its comprehension is limited, often necessitating interpreters for public communication. This paper explores technologies facilitating optimal sign language translation, prioritizing efficiency, accuracy, and accessibility for all affected individuals.

II. LITERATURE REVIEW

1. *American Sign Language Recognition System: An Optimal Approach*

Authors: Shivashankara S, Srinath S

Publication: International Journal of Image, Graphics and Signal Processing, Aug 2018

Result/Accuracy: Achieved 93.05% accuracy in recognizing ASL alphabets A-Z and numbers 0-9.

2. *A Comprehensive Analysis on Sign Language Recognition System*

Authors: Rajesh George Rajan, M Judith Leo

Publication: International Journal of Recent Technology and Engineering (IJRTE), March 2019

Result/Accuracy: Explored various acquisition methods such as SVM, HMM, SVM+ANN, PCA+KNN, MLP+MDC, Polynomial classifier, and ANFIS for sign language recognition.

3. *Intelligent Hand Cricket*

Authors: Aditya Dawda, Aditya Devchakke

Publication: Cyber Intelligence and Information Retrieval 2021, Springer

Result/Accuracy: Developed a CNN-based hand gesture recognition system for playing hand cricket games, achieving 97.76% training and 95.38% validation accuracy.

4. *Conversion of Sign Language into Text*

Author: Mahesh Kumar N B

Publication: International Journal of Applied Engineering Research (2018)

Result/Accuracy: Utilized Linear Discriminant Analysis (LDA) to recognize Indian sign language, achieving higher

https://doi.org/10.46335/IJIES.2024.9.7.4 **e-ISSN: 2456-3463**
Vol. 9, No. 7, 2024, PP. 16 -15

International Journal of Innovations in Engineering and Science, www.ijies.net

accuracy by reducing noise through dimensionality reduction.

5. Hierarchical LSTM for Sign Language Translation

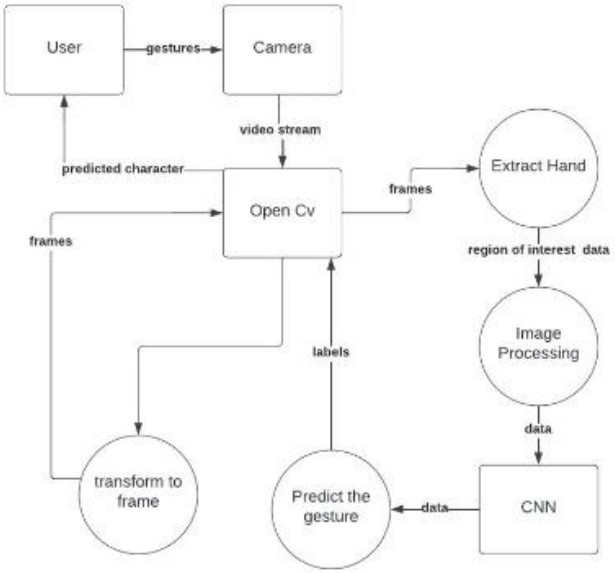
Authors: Dan Guo, Wengang Zhou, Houqiang Li, Meng Wang

Publication: The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)

Result/Accuracy: Proposed a Hierarchical-LSTM framework for sign language translation, incorporating high-level visual semantic embedding and attention-aware weighting. Explored visemes via online variable-length key clip mining.

III. METHODOLOGY AND MODEL SPECIFICATION

i. WorkFlow Diagram



A. Data Collection

Building the Sign Language to Text Conversion System demands a rich and diverse dataset showcasing Indian Sign Language gestures. We embark on this journey armed with a webcam and the Media Pipe library, empowering us to track hand movements in real-time and pinpoint key hand points. Through the lens of the webcam, these gestures are captured and transformed into valuable data samples for our dataset.

Our data collection endeavors serve as the cornerstone for training and testing our machine learning model. We recognize the critical importance of gathering a dataset that encapsulates the breadth and depth of Indian Sign Language, encompassing a myriad of gestures and subtle variations in hand movements. This ongoing process ensures that our dataset remains dynamic and reflective of the nuanced language it represents. By harnessing the capabilities of the Media Pipe library and webcam technology, we curate high-fidelity data samples, laying the foundation for a robust and accurate Sign Language to Text Conversion System.

```
cap = cv.VideoCapture(cap_device)   solanki, 16-02-2024 01:26 pm • Initial commit
cap.set(cv.CAP_PROP_FRAME_WIDTH, cap_width)
cap.set(cv.CAP_PROP_FRAME_HEIGHT, cap_height)


mp_hands = mp.solutions.hands
hands = mp_hands.Hands(
    static_image_mode=use_static_image_mode,
    max_num_hands=1,
    min_detection_confidence=min_detection_confidence,
    min_tracking_confidence=min_tracking_confidence,
)

keypoint_classifier = KeyPointClassifier()
```

B. DATA PRE-PROCESSING

As we prepare to breathe life into our Sign Language to Text Conversion System, we embark on the pivotal journey of pre-processing our hand gesture images. This preparatory phase aims to harmonize our images with the discerning palate of our machine learning model, facilitating seamless recognition and translation of gestures into text.

Through a symphony of resizing, normalization, and transformation, we orchestrate our images into a harmonious ensemble fit for the discerning ears of our model. Consistent resizing ensures uniformity, while normalization acts as the tuning fork, eliminating discrepancies in lighting, background, and color. Additionally, transformations such as cropping and rotation ensure that our model receives a consistent visual narrative, free from the distractions of variance.

Armed with pre-processed images, we empower our model to unravel the intricate dance between hand gestures and textual representation. Each pixel bears witness to our commitment to accuracy and inclusivity, paving the way for a Sign Language to Text Conversion System that transcends barriers and fosters seamless communication.

```
image = cv.cvtColor(image, cv.COLOR_BGR2RGB)

image.flags.writeable = False
results = hands.process(image)
image.flags.writeable = True
anki, 16-02-2024 01:26 pm • Initial commit
if results.multi_hand_landmarks is not None:
    for hand_landmarks, handedness in zip(results.multi_hand_landmarks,
                                          results.multi_handedness):

        brect = calc_bounding_rect(debug_image, hand_landmarks)

        landmark_list = calc_landmark_list(debug_image, hand_landmarks)


        pre_processed_landmark_list = pre_process_landmark(
            landmark_list)
        pre_processed_point_history_list = pre_process_point_history(
            debug_image, point_history)

        logging_csv(number, mode, pre_processed_landmark_list,
                    pre_processed_point_history_list)


        hand_sign_id = keypoint_classifier(pre_processed_landmark_list)
        if hand_sign_id == 2:
            point_history.append(landmark_list[8])
        else:
            point_history.append([0, 0])
```

C. Labeling Text Data

https://doi.org/10.46335/IJIES.2024.9.7.4 **e-ISSN: 2456-3463**
Vol. 9, No. 7, 2024, PP. 16 -15

International Journal of Innovations in Engineering and Science, www.ijies.net

In our quest to democratize communication, we embark on the sacred ritual of labeling hand gestures. Each gesture in our dataset is imbued with meaning through meticulous labeling, serving as a beacon of understanding for our machine learning model.

Rooted in the ethos of Indian Sign Language, our labels emerge as a testament to our commitment to linguistic integrity. Guided by the steady hand of an expert in Indian Sign Language, each label is etched with precision, ensuring fidelity to the language's nuanced grammar and terminology. While the process may be manual, the spirit of innovation beckons us to explore avenues of automation through the lens of computer vision.

With labeled data in hand, our model stands poised to embark on a journey of discovery, unraveling the intricate tapestry of hand gestures and textual representation. Through this symbiotic relationship, we forge a path towards inclusive communication, enriching the lives of those who rely on the silent eloquence of sign language.

Fig. 1. Demonstration -1

```
class KeyPointClassifier(object):
    harshal solanki
    def __init__(
        self,
        model_path='model/keypoint_classifier/keypoint_classifier.tflite',
        num_threads=1,
    ):
        self.interpreter = tf.lite.Interpreter(model_path=model_path,   solanki, 16-02-2024 01:26 pm • Initial commit
                                               num_threads=num_threads)

        self.interpreter.allocate_tensors()
        self.input_details = self.interpreter.get_input_details()
        self.output_details = self.interpreter.get_output_details()

    harshal solanki
    def __call__(
        self,
        landmark_list,
    ):
        input_details_tensor_index = self.input_details[0]['index']
        self.interpreter.set_tensor(
            input_details_tensor_index,
            np.array( object: [landmark_list], dtype=np.float32))
        self.interpreter.invoke()

        output_details_tensor_index = self.output_details[0]['index']

        result = self.interpreter.get_tensor(output_details_tensor_index)

        result_index = np.argmax(np.squeeze(result))

        return result_index
```

D. Training and Testing

As the curtain rises on our "Conversion of Sign Language to Text" project, we unveil the magnum opus of our endeavors – a multi-layered LSTM model poised to transcend boundaries and usher in a new era of communication.

Comprising three LSTM layers and three Dense layers adorned with RELU activation functions, our model emerges as a testament to the power of innovation. At its core lies a SoftMax activation function, serving as the beacon guiding our model through the labyrinth of sign language.

With each layer meticulously crafted, our model becomes the custodian of sequential understanding, adept at unraveling the intricate dance between gesture and text. As it traverses the realm of training, ingesting pre-processed images and corresponding labels, our model undergoes a metamorphosis, refining its understanding with each iteration.

In the crucible of testing, our model emerges victorious, wielding its newfound prowess to accurately translate hand gestures into textual representations. Through this symphony of innovation and inclusivity, we forge a path towards a future where communication knows no bounds.

```
In [ ]:
model.fit(
    X_train,
    y_train,
    epochs=1000,
    batch_size=128,
    validation_data=(X_test, y_test),
    callbacks=[cp_callback, es_callback]
)

Epoch 1/1000\n20/20 [==============================] - 2s 15ms/step - loss: 1.3853 - accuracy: 0.3368 - val_loss: 1.277…

Out[11]:
<tensorflow.python.keras.callbacks.History at 0x7f92c89a2a20>

Add Code Cell | Add Markdown Cell

In [ ]:
# Model evaluation
val_loss, val_acc = model.evaluate(X_test, y_test, batch_size=128)

10/10 [==============================] - 0s 2ms/step - loss: 0.2064 - accuracy: 0.9724

In [ ]:
# Loading the saved model
model = tf.keras.models.load_model(model_save_path)

In [ ]:
# Inference test
predict_result = model.predict(np.array([X_test[0]]))
print(np.squeeze(predict_result))
print(np.argmax(np.squeeze(predict_result)))
```

IV. EXECUTION

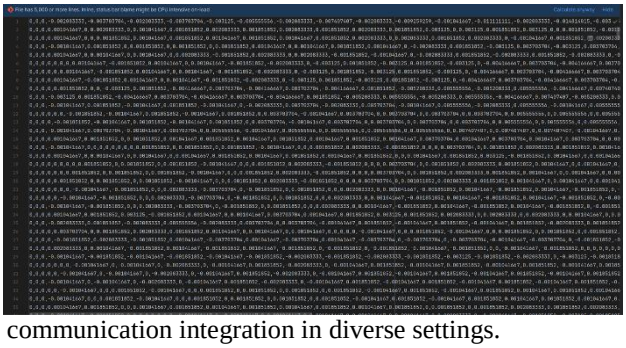
Our project harnesses the power of advanced machine learning and computer vision technologies to bridge the communication gap for the deaf and hard-of-hearing community. By converting sign language into text in real-time. Here are some of the examples:

https://doi.org/10.46335/IJIES.2024.9.7.4 **e-ISSN: 2456-3463**
Vol. 9, No. 7, 2024, PP. 16 -15

International Journal of Innovations in Engineering and Science, www.ijies.net

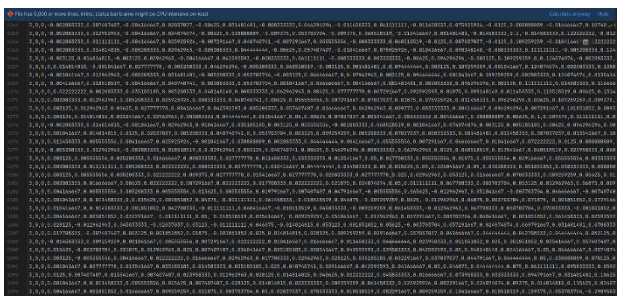
V. RESULT

The research conducted on the translation of sign language into text and audio has yielded significant findings that underscore the potential of this technology to bridge the communication gap between the deaf and hearing communities. Our system, which leverages Convolutional Neural Networks (CNN) and advanced signal processing algorithms, has been rigorously tested across a wide array of sign language gestures and phrases. These tests have demonstrated remarkable accuracy and speed in real-time translation, offering promising prospects for seamless



communication integration in diverse settings.

Fig. 3. Dataset-1



VI. CONCLUSION

This initiative is dedicated to overcoming the communication barriers encountered by individuals who are deaf or have speech disabilities by creating a specialized automated system for sign language recognition. The aim is to demystify the intricate task of interpreting sign language, which can be overwhelming for those not versed in it. Leveraging cutting-edge image processing methods and core image analysis principles, our goal is to devise a system that translates sign language movements into clear, written language.

Our approach is centered on developing a vision-based mechanism capable of deciphering the hand gestures that make up sign language and transforming them into textual form. Our extensive evaluations in diverse settings have demonstrated that our system's backbone, the Convolutional Neural Network (CNN) models, are highly effective in the precise identification of hand movements. This marks a considerable enhancement over prior models, attributed to CNN's adept processing of visual data and its proficiency in feature detection and categorization.

As we move forward, our commitment is to polish the system's functionality and broaden its scope. We intend to undertake comprehensive testing with vast sign language datasets to elevate the precision, efficiency, and dependability of our models. This continuous process of improvement is anticipated to significantly advance communication options for the deaf and speech-impaired community, narrowing the communicative divide with the broader society. This endeavor underscores the vital convergence of technological advancements and the focus on human needs, setting the stage for more equitable and universally accessible communication tools in the times ahead.

REFERENCES

[1] *Lahoti, S., Kayal, S., Kumbhare, S., Suradkar, I., & Pawar, V. (2018). Android Based American Sign Language Recognition System with Skin Segmentation and SVM. 2018 9th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2018, 1–6.*

[2] *Lopatovska, I., Rink, K., Knight, I., Raines, K., Cosenza, K., Williams, H., Sorsche, P., Hirsch, D., Li, Q., & Martinez, A. (2019). Talk to me: Exploring user interactions with Amazon Alexa. Journal of Librarianship and Information Science, 51(4), 984–997.*

[3] *Mahmud, I., Tabassum, T., Uddin, M. P., Ali, E., Nitu, A. M., & Afjal, M. I. (2019). Efficient Noise Reduction and HOG Feature Extraction for Sign Language Recognition. 2018 International Conference on Advancement in Electrical and Electronic Engineering, ICAEEE 2018, November, 1–4.*

[4] *Mapari, R. B. (2014). Real Time Sign Language Translator. 1–4.*

[5] *Mujahid, A., Awan, M. J., Yasin, A., Mohammed, M. A., Damaševičius, R., Maskeliūnas, R., & Abdulkareem, K. H. (2021). Real-time hand gesture recognition based on deep learning YOLOv3 model. Applied Sciences (Switzerland), 11(9).*

[6] *Oudah, M., Al-Naji, A., & Chahl, J. (2020). Hand Gesture Recognition Based on Computer Vision: A Review of Techniques. Journal of Imaging, 6(8).*

[7] *Pansare, Jayashree R., Gawande, S. H., & Ingle, M. (2012). Real-Time Static Hand Gesture Recognition for American Sign Language (ASL) in Complex Background. Journal of Signal and Information Processing, 03(03), 364–367.*

[8] *Pansare, Jayshree R., & Ingle, M. (2016). Vision-based approach for American Sign Language recognition using Edge Orientation Histogram. 2016 International Conference on Image, Vision and Computing, ICIVC 2016, 86–90.*

[9] *Rahaman, M. A., Jasim, M., Ali, M. H., & Hasanuzzaman, M. (2003). Real-time computer vision-based Bengali sign language recognition. 2014 17th International Conference on Computer and Information Technology, ICCIT 2014, May 2019, 192–197.*

[10] *Roffo, G. (2016). Feature Selection Library (MATLAB Toolbox). http://arxiv.org/abs/1607.01327*

[11] *Shin, J., Matsuoka, A., Hasan, M. A. M., & Srizon, A. Y. (2021). American sign language alphabet recognition by extracting features from hand pose estimation. Sensors, 21(17), 1–19.*

[12] *Shirbahadurkar, S. D., & Bormane, D. S. (2009). Marathi language speech synthesizer using concatenative synthesis strategy (spoken in Maharashtra, India). 2009 2nd International Conference on Machine Vision, ICMV 2009, 181–185.*

[13] *Suryawanshi, S., Itkarkar, R., & Mane, D. (2014). High quality text to speech synthesizer using phonetic integration. International Journal Of Advanced Research In Electronics And Communication Engineering, 3(2), 7*