SignTrack: Advancements in Real-Time Sign Language Processing for Inclusive Computing with optimized AI

Stavros Piperakis, Maria Papatsimouli, Vasilis Argyriou, *Member, IEEE,* Panagiotis Sarigiannidis, *Member, IEEE,* and George F. Fragulis, *Member, IEEE*

2024 13th International Conference on Modern Circuits and Systems Technologies (MOCAST) | 979-8-3503-8542-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/MOCAST61810.2024.10615749

***Abstract*—Visual-based translator systems, which utilize image and video data for real-time translation, represent a captivating research area. While previous research has explored text-based translation, limitations exist in capturing the nuances of human communication. Visual-based approaches address this gap by leveraging deep learning and image processing to extract information like objects, faces, and environmental context. This research project investigates the application of deep learning in sign language translation using Automated AI. By analyzing visual data, the system recognizes signs and translates them into spoken languages or generates text descriptions. This technology holds significant promise for various fields – education, entertainment, tourism, and healthcare – and can contribute to the advancement of information technology and artificial intelligence systems.**

***Index Terms*—Automated AI, IoT, Sign Language, Sign language transcription**

I. Introduction

INDIVIDUALS who are hard of hearing or imbecilic can communicate with each other, interact with others, and regularly comprehend each other by using sign language. [1]. The lives of hard of hearing are now improved regarding communication and obtaining information because of methods and software that enable sign language translation. The adoption of this technology has increased over the past ten years. Since sign language is their common language, communication between two deaf people is not difficult. However, it can be difficult for a deaf person to interact with a hearing person [2]. The recognition and translation of hand gestures using various methods have been a challenge for humanity [3].

II. Sign languages

Sign language is a form of communication used by the deaf and hard of hearing that includes facial expressions, hand gestures, and body gestures [4], [5]. Sign languages are used as spoken languages for social interaction and family communication [6] and are recognized as human languages [7], [8]. However, there are several differences between spoken and sign languages [9].

Sign languages are non-verbal languages used for communication through various hand gestures and facial expressions and for emotional expressions [10], [11]. Sign language vocabulary and syntax vary from country to country, and even from region to region within the same country. [12],[13].

Finally, sign language requires many processes, including hand configuration, movement discrimination, facial expression discrimination, and recognition of linguistically related spatial contrasts [7], [14].

A. Applications of Sign language

Interaction is one of the essential aspects of sign language software programs. Each user can become an active participant in the learning process through interaction, maintaining their attention throughout the process. Sign language is a visual language, so it can be used in any program to share information and make it easy for people who can't hear to get information. Over the years, various applications have been developed to help people learn sign languages [15] and translate signs into spoken language or text [16].

B. Approaches of sign language translation systems

Sensor-based recognition and computer vision are the two main types of technologies that sign language recognition systems often use. Due to technical image processing, a camera is used to take pictures or record videos, such as hand motion recordings. A higher-resolution camera needs more memory space and more processing power. However, computer vision systems require more expensive sensors and a lot of high-performance techniques, increasing the cost and complexity of the application. The system must maintain a factual, noise-free, and disturbance-free background [17]. Additionally, sign language recognition software is divided into [18]:

- data-gloves approaches and
- visual-based approaches.

Stavros Piperakis, Maria Papatsimouli, Panagiotis Sarigiannidis and George F. Fragulis are with the Department of Electrical and Computer Engineering, University of Western Macedonia, ZEP Campus,50100, Kozani, Greece, e-mail:ece02228@uowm.gr, maria.papatsimoulh@gmail.com, psarigiannidis@uowm.gr, gfragulis@uowm.gr.

Vasilis Argyriou is with the Department of Networks and Digital Media, Kingston University, UK, e-mail:vasileios.argyriou@kingston.ac.uk

C. Visual based approach

In the visual approach, a camera captures images and videos that are used for translation. The main advantage of visual-based methodologies is the flexibility of the framework for facial expressions, head movements, lip reading, etc. These systems are categorized into two different approaches based on how they work: the hand-crafted shading gloves and the light of skin-colour recognition. In the hand-crafted shading gloves procedure, the signer uses a colour-coded glove [19], [20]. Through colour segmentation, the colours reveal information from the image acquisition. These gloves are less expensive [18] than smart gloves and are a typical pair with unique shading on each finger and palm. A webcam and affordable colour-shaded gloves are the only required equipment. Webcams are used to capture RGB-shaded images and videos. The skin-colour technique is the most popular because signers can interact with the system directly, and all that is required is a camera to acquire the data. Additionally, the signer must perform a stop sign because the image collection process continues [20]. The limitations of smart gloves have encouraged the use of the vision-based method. In this approach, signs are captured using a camera to be recognized. The system is flexible when using the vision-based approach since these systems can be modified to incorporate non-manual signals like lip-reading, head movement recognition, and facial expressions, but there may also be issues with noise in the data processing methods [21].

D. Automated AI translation for sign languages

Sign languages are the primary languages of many people around the world. In order to overcome the communication barriers between deaf and hearing people, artificial intelligence techniques have been used with the aim of developing systems for automatic recognition and generation of sign languages. The peculiarities of signed languages must be taken into account because even though they share some features with spoken languages, they differ in other ways [22]. Till today, various automated AI translation techniques were developed for sign language translation such as [23].

III. SignTrack

SignTrack is a sign language transcriber that analyses, processes, and recognizes sign language in real-time with remarkable accuracy and efficiency. It takes a people-centered approach to computing, aiming to make computers more accessible and inclusive for everyone. By seamlessly converting sign language into written language, SignTrack breaks down communication barriers and enables meaningful connections.

A. How it works

SignTrack utilizes an LSTM model that predicts based on a sequence of data, enabling the detection of whole phrases and moving signs. To further improve precision and efficiency, SignTrack has been trained only on key hand and pose landmarks that have been extracted using MediaPipe, to save resources and offer the same accuracy on most skin shades.

B. Showcase

1) Data Collection: Forming a friendly data-collecting experience has been one of our top priorities. We developed a data collection user interface that is easy to use, even for those with minimal coding skills. Data plays a fundamental role in training an accurate and efficient model. With breaks between training sessions and intuitive design, data collection has been designed to aid our users. To further improve the ease of use, data cleaning and data augmentation are included in the data collection process (Fig. 1) . The user is warned about the conditions of data collection, determined by verifying the appropriate number of frames collected in each sequence are capturing the hands in the scene. The key landmarks of flipped images are also stored in the dataset to generate a model that can make equally accurate predictions on both hands.

Fig. 1. Successfully collecting data for the phrase ‘Thank you’

Privacy is at the centre of SignTrack. The collected data is free of personal data, like raw images. It only stores numerical values of the key-point positions as NumPy arrays, making users feel more comfortable exchanging datasets (Fig. 2)

2) Model Training: As the predictions are made based on a sequence of data, LSTM models are the most appropriate due to their ability to detect patterns within those sequences. Layers are stacked by creating a sequential model to improve the model’s ability to learn data representations (Fig. 3). The training of neural networks requires overcoming challenges and making training parameter settings that may be overwhelming for many people, introducing another barrier to

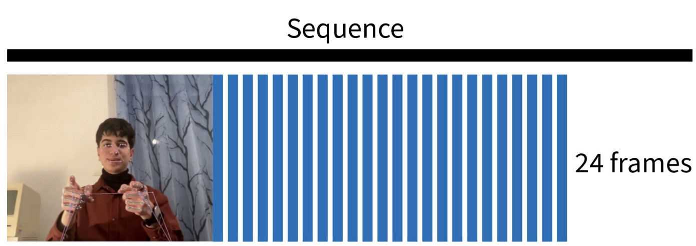

Fig. 2. The structure of collected data

accessing custom sign language recognition models. We aim to minimize such barriers by introducing AutoTrain.

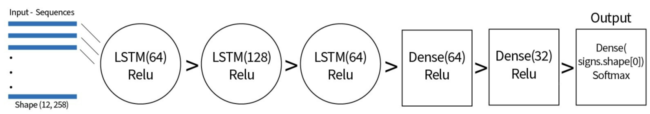

Fig. 3. The architecture of the model used

AutoTrain streamlines the training process by selecting the number of epochs and the model shape. It continuously monitors for overtraining, reverting to a previous model version if needed to prevent overfitting. Additionally, it archives the most accurate models and halts training upon reaching a preset of 150 epochs. The shape of the input and the dropout layer is also adjusted according to the structure and number of sequences collected and the number of signs, without the need of additional configuration. To overcome traditional limitations of LSTM-based models in sign language recognition from various angles, SignTrack introduces the InSpace Engine . The InSpace Engine rotates, utilizing the 3x3 Rotation Matrix, each collected point to simulate real-world signing conditions (Fig. 4).

```
# Rotate around x-axis
y_prime = y * np.cos(rx) - z * np.sin(rx)
z_prime = y * np.sin(rx) + z * np.cos(rx)

# Rotate around y-axis
x_prime = x * np.cos(ry) + z_prime * np.sin(ry)
z_prime = -x * np.sin(ry) + z_prime * np.cos(ry)

# Rotate around z-axis
x_prime = x_prime * np.cos(rz) - y_prime * np.sin(rz)
y_prime = x_prime * np.sin(rz) + y_prime * np.cos(rz)
```

Fig. 4. Code snippet from the trigonometric equations used in the InSpace Engine, derived from the 3*3 Rotation Matrix

3) SignTrack Main: Utilizing the created model turned out to be an equally fundamental part of the project. Some people may sign faster than the 12 frames that the model requires for making predictions. FastTrack is built into the SignTrack to solve this problem. It randomly appends in missing frames as explained in model training.

Optimizations form an uninterrupted experience. SignTrack uses resources only when needed. The model is called to make predictions only when the hands have been visible on the scene. While the needed punctuation is predicted after the user has completed forming the sentence (Figures 5 and 6).

4) Dataset: The dataset consists of sequences of 24 frames that form recognizable signs, based on which the model will be trained on. The collected data is free of personal data, like raw images, each sequence contains the key points extracted using Mediapipe, as NumPy arrays.

5) Dependances: This project has been developed using: Python: 3.7 Tensorflow: 2.5 OpenCV: 4.1.2.30 Scikit-Learn Matplotlib Mediapipe Cvzone Sentencepiece Transformers

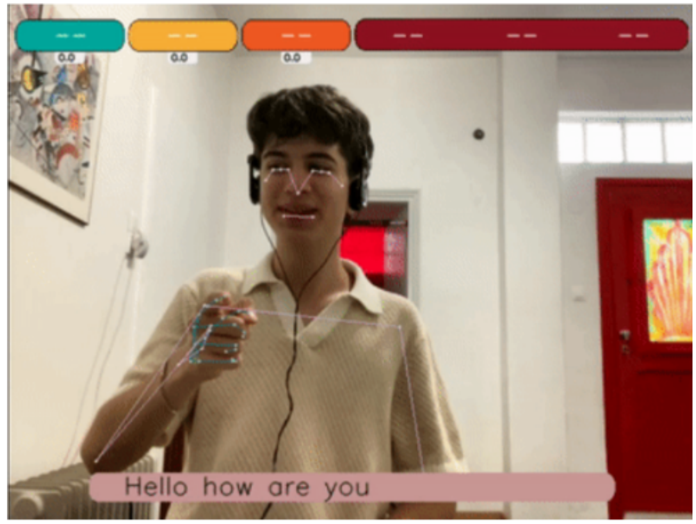

Fig. 5. Real-time transcription example 1, the displayed phrase is "Hello how are you", punctuation is added after completing a sentence to save on resources

IV. Conclusions

SignTrack, a novel sign language recognition system, promotes inclusivity through real-time translation. It uses advanced deep learning models for accurate gesture recognition and utilizes user-friendly data collection tools for building diverse datasets. SignTrack also exhibits significant performance improvements, including faster predictions and seamless word transitions. However, limitations like limited vocabulary and dependence on camera angles highlight areas for future expansions. The development of SignTrack encounters various limitations. Its performance is vulnerable to external factors, such as fluctuations in camera angle and environmental lighting. Moreover, vocabulary expansions and adaptations to regional vocabularies necessitate investment in data collection and model training, potentially causing disruptions within more localized or less populous communities. SignTrack is

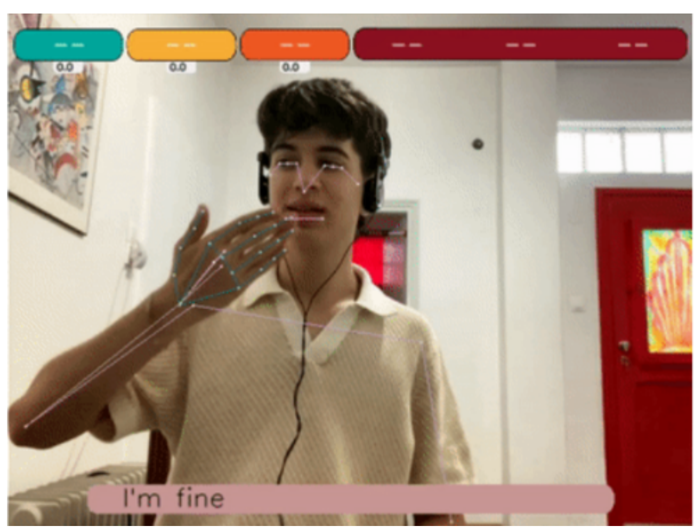

Fig. 6. Real-time transcription example 2, the displayed phrase is "I'm fine"

free and open source, ensuring the privacy and flexibility of its applications. Customers will be able to purchase specialized SignTrack recognition models covering specific languages and vocabulary. In addition, due to the grammatical differences between sign language and written language, it is necessary to use Gen AI to produce understandable and comprehensive translations. Upon completion of each sentence SignTrack will call cloud based API, for the use of which a small charge will be made on commercial uses, which will reshape the sentences accordingly. Overall, SignTrack presents a promising foundation for breaking down communication barriers and empowering the deaf and hard-of-hearing communities.

A. Future Directions

While SignTrack offers a robust foundation, future research can explore expanding the supported sign language vocabulary and incorporating advanced natural language processing techniques to capture the nuances of sign language grammar. Additionally, investigating deep learning architectures beyond LSTMs for even greater recognition accuracy holds promise.

Acknowledgment

This research received funding from the European Union's Horizon Europe Research and Innovation programme under Grant Agreement No. 101135800 (RAIDO). Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

References

[1] P. Dubey, "Sign language conversion flex sensor based on iot," *International Journal of Research in Engineeringand Science (IJRES)*, vol. 9, no. 2, pp. 69–71, 2021.

[2] D. Sturman and D. Zeltzer, "A survey of glove-based input," *IEEE Computer Graphics and Applications*, vol. 14, no. 1, pp. 30–39, 1994.

[3] Y. Wu and T. Huang, "Vision-based gesture recognition: A review. gesture-based communication in human-computer interaction,," pp. 103–115, 1999.

[4] P. S. Pooja Dubey, "Iot based sign language conversion," *International Journal of Research in Engineering and Science (IJRES)*, vol. 9, pp. 84–89, 2021. [Online]. Available: www.ijres.org

[5] R. Sutton-Spence and B. Woll, *Linguistics and sign linguistics*. Cambridge University Press, 1999, p. 1–21.

[6] K. Snoddon, "Wendy sandler and diane lillo-martin, sign language and linguistic universals. cambridge: Cambridge university press, 2006. pp. xxi, 547. pb $45.00." *Language in Society*, vol. 37, 10 2008.

[7] Z. Maalej, "Book review: Language, cognition, and the brain: Insights from sign language research," *Linguist List*, vol. http://www.linguistlist.org/issues/13/13-1631.html, 01 2002.

[8] C. Monikowski, "Language, cognition, and the brain: Insights from sign language research," *Studies in Second Language Acquisition*, vol. 26, no. 3, p. 497–498, 2004.

[9] B. T. Tervoort, "Sign language: the study of deaf people and their language: J.g. kyle and b. woll, cambridge, cambridge university press, 1985. isbn 521 26075. ix+318 pp," *Lingua*, vol. 70, no. 2, pp. 205–212, 1986. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0024384186900422

[10] P. Ambavane, R. Karjavkar, H. Pathare, S. Relekar, B. Alte, and N. K. Sharma, "A novel communication system for deaf and dumb people using gesture," in *ITM Web of Conferences*, vol. 32. EDP Sciences, 2020, p. 02003.

[11] A. Das, L. Yadav, M. Singhal, R. Sachan, H. Goyal, K. Taparia, R. Gulati, A. Singh, and G. Trivedi, "Smart glove for sign language communications," in *2016 International Conference on Accessibility to Digital World (ICADW)*, 2016, pp. 27–31.

[12] W. C. J. Stokoe, "Sign language structure: an outline of the visual communication systems of the american deaf. 1960." *Journal of deaf studies and deaf education*, vol. 10, pp. 3–37, Winter 2005.

[13] J. B. C. Christopoulos, "Sign language," *Journal ofCommunication Disorders 1*, vol. 18, no. 1-20, 1985.

[14] K. Emmorey, *Language, cognition, and the brain: Insights from sign language research*. Lawrence Erlbaum Associates Publishers, 2002.

[15] M. Papatsimouli, L. Lazaridis, K.-F. Kollias, I. Skordas, and G. F. Fragulis, "Speak with signs: Active learning platform for greek sign language, english sign language, and their translation," in *The 3rd ETLTC International Conference on Information and Communications Technology (ETLTC2021)*, vol. 102, no. 01008. SHS Web of Conferences, 2021.

[16] M. Papatsimouli, P. Sarigiannidis, and G. F. Fragulis, "A survey of advancements in real-time sign language translators: Integration with iot technology," *Technologies*, vol. 11, no. 4, p. 83, 2023.

[17] R. Wijayawickrama, T. P. Ravini Premachandra, and A. Chanaka, "Iot based sign language recognition system," *Global Journal of Computer Science and Technology*, 2021.

[18] M. Papatsimouli, K.-F. Kollias, L. Lazaridis, G. Maraslidis, H. Michailidis, P. Sarigiannidis, and G. F. Fragulis, "Real time sign language translation systems: A review study," in *2022 11th International Conference on Modern Circuits and Systems Technologies (MOCAST)*. IEEE, 2022, pp. 1–4.

[19] R. Akmeliawati, M. P.-L. Ooi, and Y. C. Kuang, "Real-Time Malaysian Sign Language Translation using Colour Segmentation and Neural Network," in *2007 IEEE Instrumentation & Measurement Technology Conference IMTC 2007*. Warsaw, Poland: IEEE, May 2007, pp. 1–6, iSSN: 1091-5281. [Online]. Available: http://ieeexplore.ieee.org/document/4258110/

[20] A. Z. Shukor, M. F. Miskon, M. H. Jamaluddin, F. bin Ali@Ibrahim, M. F. Asyraf, and M. B. bin Bahar, "A new data glove approach for malaysian sign language detection," *Procedia Computer Science*, vol. 76, pp. 60–67, 2015, 2015 IEEE International Symposium on Robotics and Intelligent Sensors (IEEE IRIS2015). [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050915037771

[21] M. C. Shubankar, B. and M. Priyaadharshini, "Iot device for disabled people," *Procedia Computer Science*, vol. 165, pp. 189–195, 2019.

[22] M. S. Buafra, "The automation of translation between sign language variants through artificial intelligence," Ph.D. dissertation, 2023.

[23] L. Baumgärtner, S. Jauss, J. Maucher, and G. Zimmermann, "Automated sign language translation: The role of artificial intelligence now and in the future." in *CHIRA*, 2020, pp. 170–177.