A Survey on Sign Language Recognition

Ananya Jain

Chattahoochee High School, 5230 Taylor Rd, Johns Creek, GA 30022, United States.
Email ID: ananyajain0719@gmail.com

---***---

Abstract – *Sign language plays a crucial role in facilitating communication for the deaf and hearing-impaired people in society. According to the World Health Organization, around 466 million people rely on sign language for approximately 300 distinct languages, while 56% of sign language remains dominated by English. Thus, developing accurate sign language recognition (SLR) models remains pivotal for a large section of society. In this work, we present a comprehensive overview of the various algorithms developed for sign language recognition. We broadly classify the different types of SLR based on the number of hands employed for hand gestures, vaying types of inputs for SLR models, etc. Further, we present a detailed examination of sensor-based SLR models that employ external objects like gloves, and vision-based SLR models that focus only on hand gestures images and videos. Recently, artificial intelligence has appeared as a promising approach for dramatically improving the accuracy of SLR models. Thus, we present a critical analysis of machine learning models employed for SLR, such as k-nearest neighbor, support vector machines, etc. Moreover, deep learning has enabled the prediction of sign languages with very high accuracy surpassing all the state-of-the-art methods. Thus, we present a detailed analysis of the state-of-the-art deep learning models employed for SLR, such as convolutional neural network and long short-term memory models.*

***Key Words*: Sign language recognition, Gestures, Machine learning, Deep learning, and Artificial intelligence.**

1.INTRODUCTION

The World Health Organisation estimates that 466 million people worldwide are deaf, including 432 million adults and 34 million children. Sign Language enables the deaf community to communicate with one another and with the outside world. Sign language assists in breaking down barriers between the deaf population and the rest of the world [1]. There are around 300 distinct sign languages in use across the world, which vary by country. In Fig. 1, we depict the global popularity of several sign languages (in percentages).

Recently, the advent of voice-activated personal digital assistants (PDAs) such as Apple's Siri and OK Google has revolutionized the way people live. However, not many changes have been made to support individuals with hearing impairments, so these technological advances are not readily available to them. In order to help translate hand gestures

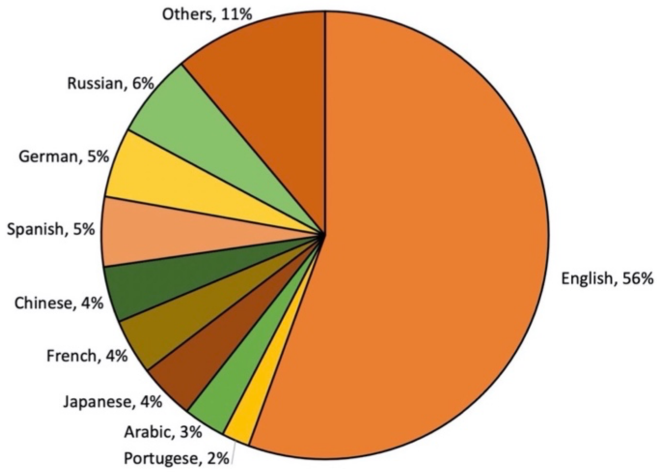

Fig-1: The global distribution of various sign languages.

into computer commands or just plain text that other people can comprehend, modern devices are either camera-compatible or include built-in cameras. However, since speech recognition systems lag sign language recognition (SLR) systems by about 30 years, SLR system development is essential to ensure that the public welfare benefits from the technological advancements even for those in the hearing-impaired community.

In recent years, one of the most promising approaches has emerged: artificial intelligence (AI). Due to significant technological advancements, artificial intelligence (AI) has drawn countless scientific and practical applications that are transforming every sector. Fig. 2 enumerates a few domains where AI is operating. One such application that has attracted a lot of interest in the last 10 years is computer vision (CV). With the use of machine learning (ML) and deep learning (DL) models, CV processes digital photos and videos as input and produces meaningful conclusions. Throughout the world, computer vision is quickly changing a wide range of sectors, including robotics, finance, agriculture, healthcare, auto-mobiles (self-driving autonomous automobiles), security systems, and many more [2]. Real-time, high-speed computing resources with minimal power consumption are necessary for the methods used to tackle CV challenges.

In this paper, we address the crucial computer vision problem of sign language recognition (SLR), which affects a significant portion of the general population. We perform a thorough literature assessment of the most recent state-of-the-art frameworks for sign language recognition (SLR). We employ deep learning (DL) and machine learning (ML) methods for examining vision-based approaches. Through presenting a summary of the state of SLR today, this paper hopes to offer insightful perspectives into this developing subject and promote developments that improve the community's ability to communicate with deaf and hearing-impaired people.

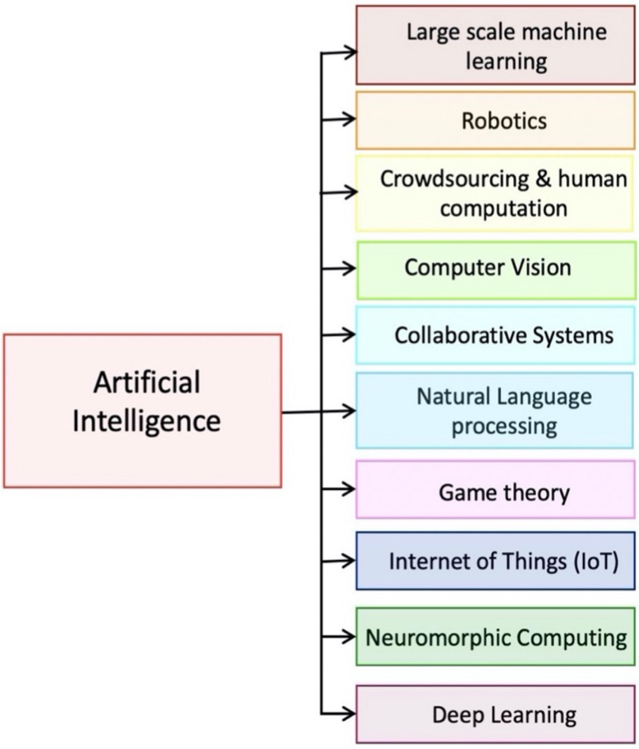

Fig-2: An example of the various domains of AI research.

2 Types of Sign Language Recognition

The signs or gestures used in sign language recognition (SLR) can be generally divided into two categories: single-handed and double-handed gestures [3], as shown in Fig. 3 and detailed below:

- One-Handed Gestures – The signer makes use of their primary hand in this instance. Both moving and static signals can be used to express it.
- Double-Handed Gestures – In this instance, the signer expresses the sign gestures with both of their hands. The Type-0 sign is made when the signer's both hands are active, while the Type-1 sign is made when their dominant hand stays more active than the non-dominant hand.

Additionally, as explained below, we can split the SLR into two main categories:

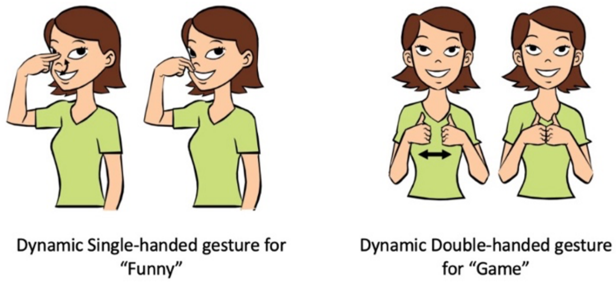

Fig-3: An example of one- and double-handed gestures.

- Isolated SLR – The fundamental task here is still the same as picture categorization issues. Since SLRs typically receive their image inputs from real-world users, they are susceptible to lighting variances, skin tone variations, distinct backdrops, and occasionally even cropped photographs. making the process of classifying them laborious. Furthermore, gathering enough data to train deep models and provide a reliable model for isolated SLR is not always feasible.
- Continuous SLR – It is a more advanced version of isolated sign language recognition, typically used for video datasets. Continuous or Video SLR not only classifies each of the video frames but also accurately places the associated labels from the supplied video segments. Generally speaking, there are two key performance evaluation criteria for video SLR. These components include consecutive alignment to produce coherent words or phrases from every input video frame, as well as the extraction of features and classification. The rapid growth of deep learning (DL) technologies, including natural language processing, long short-term memory (LSTM) models, graph neural network models, and others, has increased the quality of continuous SLR [1].

3. Sensor-based Sign Language Recognition

The field of hand and gesture recognition was introduced in 1987 by Zimmerman et al. [4], who presented a hand gesture interface that used a wearable glove's magnetic field to determine the hand's location and direction. The way the glove sensor worked was by identifying different hand gestures or finger motions and then relaying the information to other devices for classification and processing. To extract the hand's characteristics, Oz and Leu [5] integrated the principles of artificial intelligence with sensor gloves, commonly known as Cyber gloves, and a 3-D movement tracker. Then, using artificial neural networks (ANNs), a model was trained on these variables to predict American Sign Language (ASL). The accuracy provided by the suggested model was 95%. There were drawbacks to this

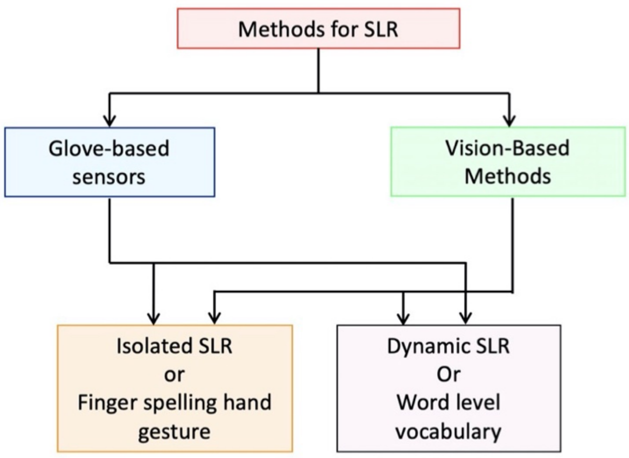

Fig-4: An example of several SLR approaches.

strategy, too, including the signer's perception of wearing the gloves as constricted and strange, and the expensive cost of the wearable gloves made it impractical for everyone to buy them [1]. We describe the various SLR techniques comprehensively in Fig. 4 for clarity.

4. Vision-based Sign Language Recognition

With the introduction of AI and significant technological developments, vision-based techniques were introduced to automate the SLR. Because only one camera is needed, vision-based methods are more affordable and user-friendly than sensor-based ones while providing signers with greater normalcy and mobility [1]. In order to apply AI paradigms for SLR, cameras or other devices capture images of hand gestures, from which key information such as hand location, finger position, posture, and facial expression are extracted. The ML and DL algorithms are then given the extracted features to classify different gestures.

5. Machine Learning-based Sign Language Recognition

A subset of artificial intelligence (AI), machine learning (ML) involves the creation of algorithms that mimic computers' ability to learn novel abilities or expertise from previously acquired information or data while also constantly enhancing their own performance. We describe some of the popular machine learning techniques in Fig. 5. Statistics, likelihood theory, approximation theory, geometric analysis, and algorithmic complexity theory are all included in the interdisciplinary field of machine learning [6]. We will go into more detail regarding the ML-based SLR models in this subsection, as shown in Fig. 5 below:

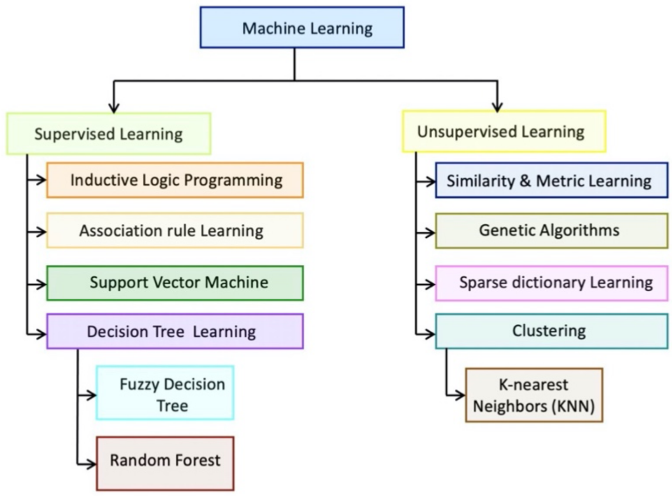

Fig-5: Common ML techniques used in SLR.

- Support Vector Machine (SVM)-based SLR model – Perhaps the most prominent ML methods for supervised learning for SLR is the utilization of SVMs. It is a margin-based non-linear regression technique with excellent generalizability, traceability, and high accuracy capabilities, making it a good fit for classification tasks. Kumar et al. [7] proposed a model for ASL using SVM. The suggested approach used manual feature extraction to classify both static and dynamic indicators. For both single- and double-handed Indian Sign Language (ISL) signs, Athira et al. [8] suggested a different approach based on multi-class SVM. SVM models performed superior for static signs over for dynamic signs, as seen by their machine learning model's 90.1% accuracy for static signs and 89% accuracy for dynamic signs.
- K-nearest neighbors (KNN)-based SLR model – KNN is an unsupervised machine learning technique that addresses regression and classification issues. The fundamental tenet of KNN is that feature sets with comparable characteristics may correspond to similar objects. In order to propose a model to categories sign gestures using a KNN classifier, Gupta et al. [9] used features from histogram-oriented gradient (HOG) and scale invariant feature transform (SIFT). For the single-handed gesture, the accuracy of the suggested model was 97.50%, while for the double-handed gesture, it was 91.11%. The aforementioned models operate differently for different signs, even though they provide acceptable accuracy on some datasets due to the fact that their performance is dependent on the features that have been extracted.
- Decision Tree-based SLR model – Fang et al. [10] developed an SLR model for an extensive vocabulary set utilizing a fuzzy Decision Tree in

order to overcome the problem with KNN models. Decision tree is a non-parametric supervised ML algorithm. The entire information is considered at the root node, which is then divided into child nodes according to decision criteria, resulting in each leaf node representing a distinct class level. By imposing a fuzzy model on DT, fuzzy DT permits each leaf node to have a partial membership. The accuracy achieved by the suggested model was 91.6%. Su et al. [11] presented a different random forests (RF)-based model for the recognition of Chinese sign language (CSL) subwords. The model classified 100 frequently used subwords with an accuracy of 98.25%.

6. Deep Learning-based Sign Language Recognition

As a result of developments in high-speed computing technology, cloud computing, and big data, machine learning (ML) has expanded to include deep learning (DL). We expand on a few of the popular deep learning techniques in Fig. 6. By incorporating several hidden layers with non-linear activation functions, DL networks are able to model far more complex cases than ML models. These layers are made up of neurons that use data-driven training to learn intricate approximations of non-linear functions. In contrast to machine learning models, which necessitate a laborious, primarily human-based feature extraction workflow, models based on DL do not require feature extraction prior to calculations [12]. We will now go into more depth about the DL-based models used for SLR, which are shown in Fig. 6 below:

- Neural Network (NN)-based SLR model – Islam et al. [13] demonstrated a real-time hand gesture recognition system using NN. Employing a camera to gather 1850 single-handed static signs, the data provided by the model was 94.32% accurate in this study. Signs were gathered from a digital camera by Karayilan and Kiliç [14], who then retrieved raw and histogram information. Their suggested neural network model produced raw feature accuracy of 70% and histogram feature accuracy of 85%.
- Convolutional Neural Network (CNN)-based SLR model – The most popular deep learning technique for extracting spatial characteristics, particularly from picture inputs (covered in section 3.5), is convolutional neural network, or CNN. It requires no pre-processing, which makes it ideal for creating SLR systems. A CNN-based model for English sign language recognition was proposed by Krishnan et al. [15]; it had three convolutional layers and three max-pooling layers, and it achieved an accuracy of 82%. Employing impulse radio sensors for American Sign Language (ASL), Kim et al. [16] created another hand motion detection model. With a median accuracy of more than 90%, they presented a model based on CNN for the classification of gathered indicators. In order to create a multimodal SLR system, Ferreira et al. [17] gathered 1400 single-handed stationary signs combining color, depth, and jump motion data. Using a CNN classifier, their suggested framework produced the best accuracy of 97%. Oyedotun and Khashman [18] suggested another CNN-based static hand identification method employing 2040 signals. After segmenting the collected indicators using median filtering, they put forth a CNN-based model that produced a 91.33% accuracy rate.
- Long Short-Term Memory (LSTM)-based SLR model – CNN-based models perform poorly for dynamic sign movements but very accurately for static ones. With an accuracy of 99%, He [19] presented a model that retrieved the features employing the colored R-CNN technique. Long Short-term Memory (LSTM) models are appropriate for dynamic sign recognition since the SLR system has to coordinate each of the labels as sequential data in the context of dynamic sign recognition and sentence level recognition [20]. Long-term information retention is a clear benefit of LSTM models, which makes them very useful [21].

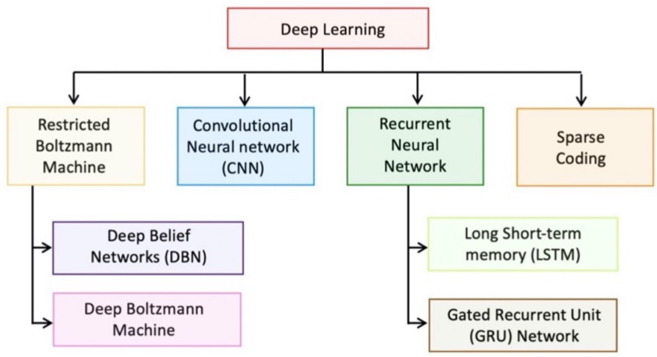

Fig-6: Common methods for DL employed in SLR.

We provide a thorough overview of current efforts in Table 1 for the purpose of clarity. In this table, we include the ML/DL algorithm, features used, dataset used, and accuracy for each of these references.

Table -1: A summary of different models for sign language recognition.

Ref No.	Features	Model	Dataset	Accuracy
[22]	-	SVM, KNN	ISL	95.3%, 89.9%
[23]	Skeletal	GCN	ASLLVD-	61.04%

	data		Skeleton	
[24]	RGB	RNN, NN	RWTH-Phoenix Weather	0.281 (Error)
[25]	Skeletal data	GCN	DHG-14/28	91.9%
[26]	RGB, Skeletal data	RNN	GMU-ASL-51	95%
[27]	RGB, Depth motion	CNN	FASL-RGB	95.62%
[28]	Dynamic and RGB	CNN	NVIDIA Benchmarks	83.83%
[29]	-	SVM	ISL	98.1%
[30]	Dynamic and RGB	CNN	RWTH Boston-50	89.33%
[31]	-	KNN	ISL	89.8%
[32]	Gloves	NN	-	95.2%
[33]	RGB, Depth	CNN	ASL Finger A	98.13%
[1]	Dynamic, RGB and Depth	CNN	NATOPS	95.87%
[1]	Dynamic, RGB and Depth	CNN	SBU	97.51%
[34]	Arm Sensors	Decision Tree, SVM	-	81.88%, 99.11%

7. CONCLUSIONS

In conclusion, this survey paper illuminates the crucial role of sign language recognition (SLR) in society for improving communication with the deaf and hearing-impaired people. In this survey, we have deeply analyzed different types of SLR and various methodologies to improve the SLR. Based on the number of hands employed in performing the sign language we can broadly classify the SLR as one-handed and two-handed models. Further, we can classify different algorithms developed for the SLR as isolated and continuous SLR models based on the static image and video dataset employed, respectively. Further, we investigated the sensor-based and vision-based SLR models. We show that the vision-based SLR models are superior to the sensor-based SLR models because they employ only hand gestures compared to the specialized gloves, which increases the cost and remains easy to use. Moreover, we analyze the great impact of artificial intelligence on the SLR models. We analyze the impact of various machine learning-based SLR models, such as SVM, KNN, and decision trees. We found that among other classifiers including KNN and Naive Bayes, SVM has emerged as a promising contender. However, ML models require data processing and feature extraction, which remains subjective to human knowledge and costly operation. Thus, we also perform analysis on DL models for the SLR. We prepare a detailed analysis of the various DL models employed for the SLR, such as NN, CNN, and LSTM.

REFERENCES

[1] R. Rastgoo, K. Kiani, S. Escalera, Sign Language Recognition: A Deep Survey, Expert Systems with Applications, vol. 164, 2021.

[2] J. Chai, H. Zeng, A. Li, E. W.T. Ngai, "Deep learning in computer vision: A critical review of emerging techniques and application scenarios," Machine Learning with Applications, vol. 6, 2021.

[3] A. Wadhawan and P. Kumar, "Sign language recognition systems: A decade systematic literature review," Archives of Computational Methods in Engineering, vol. 28(3), pp. 785–813, 2021.

[4] Zimmerman, Tom G. et al. "A hand gesture interface device." International Conference on Human Factors in Computing Systems, 1987.

[5] C. Oz and M. C. Leu, "Linguistic properties based on american sign language isolated word recognition with artificial neural networks using a sensory glove and motion tracker," Neurocomputing, vol. 70, pp. 2891-2901, 2007.

[6] C. M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

[7] A. Kumar, K. Thankachan, and M. M. Dominic, "Sign language recognition," in IEEE 2016 3rd International Conference on Recent Advances in Information Technology (RAIT), pp. 422–428, 2016.

[8] P. Athira, C. Sruthi, and A. Lijiya, "A signer independent sign language recognition with co-articulation elimination from live videos: an Indian scenario," Journal of King Saud University-Computer and Information Sciences, 2019.

[9] B. Gupta, P. Shukla, and A. Mittal, "K-nearest correlated neighbor classification for Indian sign language gesture recognition using feature fusion," in 2016 IEEE International conference on computer communication and informatics (ICCCI), pp. 1–5, 2016.

[10] G. Fang, W. Gao, and D. Zhao. "Large vocabulary sign language recognition based on fuzzy decision trees," IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, vol. 34(3), pp. 305–314, 2004.

[11] R. Su, X. Chen, S. Cao, and X. Zhang, "Random forest-based recognition of isolated sign language subwords using data from accelerometers and surface electromyographic sensors," Sensors, vol. 16, pp. 100, 2016.

[12] I. J. Goodfellow, Y. Bengio, and A. Courville. Deep Learning. MIT Press, Cambridge, MA, USA, 2016.

[13] M. M. Islam, S. Siddiqua, and J. Afnan, "Real-time hand gesture recognition using different algorithms based on

american sign language," in 2017 IEEE international conference on imaging, vision & pattern recognition (icIVPR), pp. 1–6, 2017.
[14] T. Karayılan and Ö. Kılıç. "Sign language recognition," in IEEE 2017 International Conference on Computer Science and Engineering (UBMK), pp. 1122–1126. 2017.
[15] P. T. Krishnan and P. Balasubramanian, "Detection of alphabets for machine translation of sign language using deep neural net," in IEEE 2019 International Conference on Data Science and Communication (IconDSC), pp. 1–3, 2019.
[16] S. Y. Kim, H. G. Han, J. W. Kim, S. Lee, and T. W. Kim, "A hand gesture recognition sensor using reflected impulses," IEEE Sensors Journal, vol. 17(10), pp. 2975–2976, 2017.
[17] P. M. Ferreira, J. S. Cardoso, and A. Rebelo. "Multimodal learning for sign language recognition," in Springer - Iberian conference on pattern recognition and image analysis, pp. 313– 321, 2017.
[18] O. K. Oyedotun and A. Khashman, "Deep learning in vision-based static hand gesture recognition," Neural Computing and Applications, vol. 28(12), pp. 3941–3951, 2017.
[19] S. He, "Research of a sign language translation system based on deep learning," in IEEE 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM), pp. 392–396, 2019.
[20] S. Das, M. Chakraborty, B. Purkayastha, et al., "A review on sign language recognition (slr) system: Ml and dl for slr." in 2021 IEEE International Conference on Intelligent Systems, Smart and Green Technologies (ICISSGT), pp. 177–182, 2021.
[21] S. Aly and W. Aly, "Deepar slr: A novel signer-independent deep learning framework for isolated arabic sign language gestures recognition," in IEEE Access, vol. 8, pp. 83199–83212, 2020.
[22] R. Agarwal, "Bayesian k-nearest neighbour-based redundancy removal and hand gesture recognition in isolated Indian sign language without materials support," In IOP Conference Series: Materials Science and Engineering, vol. 1116, pp. 012126, 2021.
[23] C. C. d. Amorim, D. Macêdo, and C. Zanchettin, "Spatial-temporal graph convolutional networks for sign language recognition," in Springer International Conference on Artificial Neural Networks, pp. 646–657, 2019.
[24] M. Borg and K. P. Camilleri, "Phonologically-meaningful subunits for deep learning-based sign language recognition,," in Springer European Conference on Computer Vision, pp. 199–217, 2020.
[25] Y. Chen, L. Zhao, X. Peng, J. Yuan, and D. N. Metaxas, "Construct dynamic graphs for hand gesture recognition via spatial-temporal attention," arXiv preprint arXiv:1907.08871, 2019.
[26] A. A. Hosain, P. S. Santhalingam, P. Pathak, H. Rangwala, and J. Kosecka. "Hand pose guided 3d pooling for word-level sign language recognition," in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 3429–3439, 2021.
[27] D. Huh, S. Gurrapu, Y. Du, J. Deorukhkar, F. Olson, H. Rangwala, P. Pathak, and J. Kosecka,, "Generative multi-stream architecture for american sign language recognition," in 2019 IEEE MIT Undergraduate Research Technology Conference (URTC), pp. 1–4, 2019.
[28] O. Köpüklü, A. Gunduz, N. Kose, and G. Rigoll, "Real-time hand gesture detection and classification using convolutional neural networks," in 2019 14th IEEE International Con- ference on Automatic Face & Gesture Recognition (FG 2019), pp. 1–8, 2019.
[29] H. Lilha and D. Shivmurthy, "Analysis of pixel level features in recognition of real life dual-handed sign language data set," in 2011 IEEE international conference on recent trends in information systems, pp. 246–251, 2011.
[30] K. M. Lim, A. W. C. Tan, C. P. Lee, and S. C. Tan, "Isolated sign language recognition using convolutional neural network hand modelling and hand energy image," Multimedia Tools and Applications, vol. 78(14), pp. 19917–19944, 2019.
[31] A. Nandy, J. S. Prasad, S. Mondal, P. Chakraborty, and G. C. Nandi, "Recognition of isolated Indian sign language gesture in real time," in Springer International conference on business administration and information processing, pp. 102–107, 2010.
[32] C. Oz and M. C. Leu, "American sign language word recognition with a sensory glove using artificial neural networks," Engineering Applications of Artificial Intelligence, vol. 24, pp. 1204– 1213, 2011.
[33] R. Rastgoo, K. Kiani, and S. Escalera, "Multi-modal deep hand sign language recognition in still images using restricted boltzmann machine," in Entropy, vol. 20, pp. 809, 2018.
[34] J. Wu, Z. Tian, L. Sun, L. Estevez, and R. Jafari, "Real-time american sign language recog- nition using wrist-worn motion and surface emg sensors," in 2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN), pp. 1–6, 2015.