

OBJECT DETECTION OF STACKED HAND GESTURES IN SIBI SIGN LANGUAGE WITH LSTM-HOLISTIC ALGORITHM

DETEKSI OBJEK BERTUMPUK GERAK TANGAN BAHASA ISYARAT SIBI DENGAN ALGORITMA LSTM-HOLISTIC

Dadang Iskandar Mulyana¹, Guruh Taruno Putra²
Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika^{1,2}
c2yp70@gmail.com¹

ABSTRACT

The challenges faced by individuals with hearing impairments in communicating, especially in environments that are less conducive to visual communication, emphasize the importance of sign language as a primary medium. In Indonesia, two prominent sign language systems are SIBI (Sistem Isyarat Bahasa Indonesia) and BISINDO (Bahasa Isyarat Indonesia). However, the complexity and nuances of sign language, including stacked gestures and dynamic gestures, pose significant barriers to accurate recognition and translation. This study proposes a real-time SIBI sign language recognition system utilizing the Long Short-Term Memory (LSTM) algorithm and the Mediapipe-Holistic framework. This technology is synchronized to process dynamic gestures and stacked hand gestures efficiently. This system is designed to translate sign language gestures into text. Using machine learning techniques, this system will use 14,400 privately collected data for each iconic language gesture. To train the model to recognize gestures and be able to adapt to various communication styles and environmental conditions, the data will be divided into 80:15. This study achieved a detection accuracy of 99.3% based on the F1-Score which allows the model to accurately detect SIBI sign language gestures

Keywords: Sign Language, Gesture Detection, Gesture Recognition, LSTM, Mediapipe.

ABSTRAK

Tantangan yang dihadapi individu dengan gangguan pendengaran dalam berkomunikasi, terutama di lingkungan yang kurang mendukung komunikasi visual, menegaskan pentingnya bahasa isyarat sebagai media utama. Di Indonesia, dua sistem bahasa isyarat yang menonjol adalah SIBI (Sistem Isyarat Bahasa Indonesia) dan BISINDO (Bahasa Isyarat Indonesia). Namun, kompleksitas dan nuansa bahasa isyarat, termasuk gerakan bertumpuk dan gerakan dinamis, menjadi hambatan signifikan dalam pengenalan dan penerjemahan yang akurat. Penelitian ini mengusulkan sistem pengenalan bahasa isyarat SIBI secara real-time dengan memanfaatkan algoritma Long Short-Term Memory (LSTM) dan kerangka Mediapipe-Holistic. Teknologi ini diintegrasikan untuk memproses gerakan dinamis dan gerakan tangan bertumpuk dengan efisien. Sistem ini dirancang untuk menerjemahkan gerakan bahasa isyarat menjadi teks. Dengan teknik pembelajaran mesin, sistem ini akan menggunakan data yang diambil secara privat sebanyak 14,400 untuk setiap gerak bahasa isyarat. untuk melatih model agar dapat mengenali gerakan serta mampu beradaptasi dengan berbagai gaya komunikasi dan kondisi lingkungan data akan dibagi menjadi 80:15. Penelitian ini mencapai akurasi deteksi 99.3% berdasarkan F1-Score yang memungkinkan model untuk melakukan deteksi gerak bahasa isyarat SIBI secara akurat.

Kata Kunci: Bahasa Isyarat, Deteksi Gerak, Pengenalan Gerak, LSTM, Mediapipe
Kata kunci: Pemantauan

PENDAHULUAN

Individu dengan gangguan pendengaran atau tunarungu seringkali menghadapi hambatan signifikan dalam interaksi sosial sehari-hari, terutama dalam lingkungan yang kurang mendukung komunikasi visual. Bahasa isyarat, sebagai sarana utama komunikasi dalam komunitas tunarungu, memiliki peran ganda: alat komunikasi yang kaya makna dan

representasi identitas budaya serta solidaritas. Namun, tantangan komunikasi tetap ada, terutama antara individu tunarungu dan mereka yang tidak memiliki gangguan pendengaran. Di Indonesia, Sistem Isyarat Bahasa Indonesia (SIBI) atau Bahasa Isyarat Indonesia (BISINDO) menjadi bahasa isyarat yang umum digunakan, dengan BISINDO sebagai bahasa yang diusulkan oleh Gerakan

Kesejahteraan Tunarungu Indonesia (GERKATIN) dan dikembangkan oleh masyarakat dengan gangguan pendengaran.

Laporan WHO (2018) mencatat peningkatan signifikan jumlah individu dengan gangguan pendengaran di Indonesia, menekankan kebutuhan mendesak akan solusi efektif untuk mengurangi kesenjangan komunikasi. Teknologi dan inisiatif inklusif menjadi kunci untuk memungkinkan individu tunarungu berkomunikasi lebih mudah dan setara, mendorong integrasi sosial yang lebih baik. Meskipun teknologi telah menawarkan solusi untuk mendeteksi bahasa isyarat, tantangan utama terletak pada keragaman, kerumitan, dan luasnya kosa kata bahasa isyarat. Perbedaan halus dalam bentuk, arah, dan gerakan setiap isyarat, ditambah variasi gaya komunikasi individu, menambah kompleksitas deteksi. Tantangan teknis juga muncul dalam mendeteksi lokasi tangan secara akurat, terutama dalam gerakan cepat atau posisi saling menutupi, yang memerlukan analisis data visual yang sangat detail.

Kamus SIBI menunjukkan karakteristik unik pada setiap bentuk isyarat, dengan beberapa isyarat memiliki kemiripan yang memerlukan akurasi tinggi dalam pengenalan. Untuk mengatasi tantangan ini, pendekatan berbasis teknologi seperti Long Short-Term Memory (LSTM) dan Mediapipe akan diterapkan. LSTM, sebagai jaringan saraf tiruan yang unggul dalam menganalisis data sekuensial, mampu memproses pola gerakan dinamis dalam bahasa isyarat. Mediapipe, di sisi lain, digunakan untuk menangkap dan melacak gerakan tangan secara real-time dengan presisi tinggi. Kombinasi ini diharapkan dapat meningkatkan efisiensi deteksi dan prediksi bahasa isyarat, mendukung inklusivitas dalam komunikasi antar individu.

Untuk mengimplementasikan sistem deteksi bahasa isyarat, pemanfaatan teknologi Machine Learning menjadi

langkah krusial. Setiap gambar atau sekuens gerakan bahasa isyarat akan dimasukkan ke dalam model pembelajaran mesin untuk dilatih, memungkinkan model mengenali dan mendeteksi makna isyarat. Pendekatan Dynamic Sign Language (DSL) akan diterapkan, melibatkan pembuatan kumpulan data gambar atau video dalam jumlah besar dengan karakteristik beragam untuk mencakup berbagai variasi gaya komunikasi. Penelitian ini bertujuan untuk mengembangkan sistem deteksi bahasa isyarat SIBI yang dirancang untuk mendukung dan meningkatkan komunikasi antara individu dengan dan tanpa gangguan pendengaran. Sistem ini diharapkan dapat meminimalkan hambatan komunikasi, memungkinkan individu dengan gangguan pendengaran menyampaikan pesan dan memahami percakapan dengan lebih mudah.

Penelitian ini mengidentifikasi masalah yang berfokus pada tantangan deteksi gerak bahasa isyarat menggunakan LSTM dan upaya peningkatan akurasi deteksi dengan Mediapipe Holistic. Beberapa masalah utama yang diidentifikasi meliputi kebutuhan data besar untuk deteksi objek menggunakan CNN tradisional dan tidak terdeteksinya objek yang hanya terlihat sebagian atau tertutup objek lain. Berdasarkan identifikasi masalah ini, rumusan masalah penelitian mencakup pertanyaan tentang peningkatan efisiensi data pada deteksi gerak bahasa isyarat bertumpuk dan cara mengatasi tantangan deteksi gerak bahasa isyarat yang hanya terlihat sebagian. Penelitian ini bertujuan untuk meneliti pengembangan efektif untuk mendeteksi gerak bahasa isyarat dengan pendekatan pada algoritma LSTM dan Mediapipe Holistic, memperkenalkan penggabungan metodologi LSTM dan Mediapipe Holistic untuk deteksi gerak bahasa isyarat yang lebih baik.

Kontribusi penelitian ini adalah sistem deteksi bahasa isyarat SIBI yang dirancang sebagai penerjemah bahasa

isyarat Indonesia menjadi bentuk alfabet atau tulisan. Sistem ini menggunakan LSTM dan Mediapipe-Holistic untuk mengenali gerakan tangan dan tubuh, mengatasi keterbatasan deteksi pada objek yang bertumpuk atau hanya terlihat sebagian. Diharapkan kontribusi ini dapat membantu orang dengan gangguan pendengaran berinteraksi dengan orang tanpa gangguan pendengaran di masa depan

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kualitatif melalui analisis literatur untuk memahami konteks dan tantangan deteksi objek, serta memilih LSTM dan Mediapipe Holistic sebagai metode yang sesuai.

Model LSTM akan dilatih menggunakan data koordinat dari Mediapipe Holistic untuk mendeteksi bahasa isyarat SIBI. Dataset yang digunakan terdiri dari 26 huruf alfabet bahasa isyarat Indonesia, dengan 22 gambar per huruf dari berbagai peraga. Untuk menghindari underfitting, ditambahkan 300 gambar baru per huruf. Setiap gambar akan diproses menggunakan Mediapipe Holistic untuk mendeteksi 21 titik landmark tangan. Data landmark akan disimpan dalam format array sebagai input model.

Metodologi penelitian mencakup identifikasi masalah (kurangnya akurasi deteksi objek bertumpuk dan sebagian terlihat), pengumpulan data (dari Kaggle dan data privat), perancangan program (arsitektur sistem dan alur kerja), implementasi algoritma (praproses data dan pelatihan LSTM), dan evaluasi hasil (presisi, recall, dan F1-score).

Rancangan pengujian akan menjawab research question dan objective penelitian. Dataset akan dibagi menjadi data pelatihan dan pengujian. Metode evaluasi mencakup akurasi deteksi (persentase objek yang terdeteksi dengan benar) dan pengujian pada berbagai

OS	Windows 11 Home 64-bit (10.0, Build 22631) (22621.ni_release.220506-1250)
CPU	AMD Ryzen 5 5500U (12 CPUs), ~2.1GHz
RAM	16GB (16384MB RAM)
GPU	Radeon Graphics
Camera	HP TrueVision HD Camera

kondisi (objek bertumpuk dan sebagian terlihat).

IMPLEMENTASI DAN PENGUJIAN

Tabel 1. Alat penelitian

Pengumpulan Data

Data yang digunakan dalam penelitian ini dibagi menjadi dua bagian: data pelatihan (85%) dan data pengujian (15%). Data pelatihan digunakan untuk melatih model LSTM agar dapat mengenali pola gerakan tangan dalam bahasa isyarat SIBI. Data pengujian digunakan untuk mengevaluasi kinerja model setelah pelatihan.

Data dikumpulkan dengan merekam video gerakan bahasa isyarat selama 480 detik per gerakan, dengan 30 frame per detik. Video kemudian diproses menggunakan Mediapipe untuk mengekstrak 258 keypoint (titik-titik penting) dari tangan dan tubuh. Pengambilan keypoint dilakukan menggunakan program yang dirancang khusus dengan kamera bawaan laptop. Variasi jarak perekaman juga dilakukan untuk meningkatkan kemampuan generalisasi model.

Model jaringan saraf tiruan yang digunakan terdiri dari beberapa lapisan LSTM dan GRU, serta lapisan Dense. Lapisan LSTM pertama menerima input 30 timesteps dan 258 fitur, dengan aktivasi ReLU dan return sequences True. Lapisan GRU ditambahkan untuk membantu LSTM dalam pengenalan gerak. Lapisan LSTM ketiga menerima output dari GRU dan LSTM sebelumnya, dengan return sequences False. Lapisan Dense digunakan untuk memproses fitur dari LSTM dan GRU. Lapisan terakhir adalah lapisan

Dense dengan aktivasi softmax untuk klasifikasi multikelas.

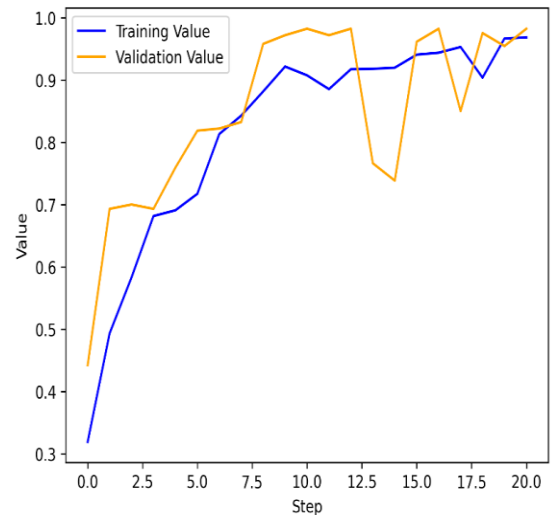
Model dilatih menggunakan data pelatihan dan dievaluasi menggunakan data pengujian. Proses pelatihan menggunakan 50 epochs dan batch size 8. Early stopping dan ReduceLROnPlateau juga digunakan untuk mencegah overfitting dan meningkatkan efisiensi pelatihan.

Hasil Penelitian

Sebagai bagian akhir dari proses penelitian, pengujian system akan dilakukan untuk mengevaluasi kinerja model yang telah dibangun dalam mendeteksi gerakan bahasa isyarat SIBI secara *real-time* menggunakan algoritma *Long Short-Term Memory (LSTM)* yang terintegrasi dengan *Mediapipe Holistic*. Berikut adalah hasil dari model yang sudah dikembangkan dapat dilihat tabel.

Tabel 2. LSTM Model Summery

Layer (type)	Output Shape	Param #
lstm_1 (LSTM)	(None, 30, 64)	82,688
gru (GRU)	(None, 30, 128)	74,496
lstm_2 (LSTM)	(None, 64)	49,408
dropout_1 (Dropout)	(None, 64)	0
dense_1 (Dense)	(None, 64)	4,160
dense_2 (Dense)	(None, 32)	2,080
dense_3 (Dense)	(None, 4)	132
Total params: 638,894 (2.44 MB)		
Trainable params: 212,964 (831.89 KB)		
Non-trainable params: 0 (0.00 B)		
Optimizer params: 425,930 (1.62 MB)		



Gambar 4.2.1 Train & Validation

Proses pengujian dilakukan dengan membagi data menjadi data pelatihan dan data pengujian sesuai dengan prosedur yang telah dijelaskan sebelumnya. Setiap gerakan direkam dengan berbagai variasi jarak dan durasi untuk memastikan kemampuan generalisasi sistem. Pengujian dilakukan dengan menggunakan metrik evaluasi F1-Score dan confusion matrix

Proses pengujian dilakukan dengan membagi data menjadi data pelatihan dan data pengujian sesuai dengan prosedur yang telah dijelaskan sebelumnya. Setiap gerakan direkam dengan berbagai variasi jarak dan durasi untuk memastikan kemampuan generalisasi sistem. Pengujian dilakukan dengan menggunakan metrik evaluasi F1-Score dan confusion matrix.

Pengujian model Long Short-Term Memory (LSTM) yang digunakan dalam penelitian ini menunjukkan hasil yang sangat memuaskan dengan nilai F1-score weighted sebesar 0.9930 (99%) dapat dilihat pada tabel

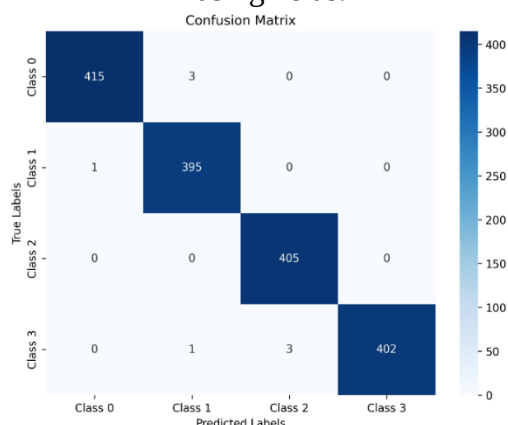
Tabel 3. Clasification

Classification Report:				
	Precision	Recall	F1-score	Support
0	1.00	1.00	1.00	60
1	1.00	1.00	1.00	82
2	1.00	0.97	0.99	73
3	0.97	1.00	0.99	72

Accuracy	0.99	287
Macro avg	0.99	287
Weighted avg	0.99	287
F1-Score (Weighted): 0.9930		

Berdasarkan hasil pengujian, model LSTM menunjukkan kemampuan yang sangat baik dalam mendeteksi gerakan bahasa isyarat SIBI dengan tingkat akurasi yang tinggi pada berbagai kelas pengujian.

Dalam pengujian confusion matrix dapat dilihat nilai diagonal pada gambar menunjukkan bahwa prediksi benar yang dilakukan oleh model untuk masing masing kelas.



Gambar 1. Confusion Matrix

Dari *confusion matrix* ini, dapat disimpulkan bahwa model memiliki performa yang sangat baik dengan jumlah kesalahan yang minimal. Tingginya jumlah nilai pada diagonal matriks menunjukkan akurasi yang tinggi dan distribusi kesalahan yang rendah di antara kelas-kelas yang diuji.

PEMBAHASAN DAN PENUTUP

Penelitian ini berhasil mengembangkan model Long Short-Term Memory (LSTM) yang mampu mendeteksi gerakan bahasa isyarat dengan akurasi yang baik. Hasil pengujian menunjukkan nilai *F1-score weighted* sebesar 99%, yang mengindikasikan keseimbangan optimal antara presisi dan *recall* dalam deteksi berbagai gerakan bahasa isyarat.

Penerapan Mediapipe Holistic untuk ekstraksi *keypoint* dari video berdurasi 480

detik terbukti efektif dalam menghasilkan data akurat untuk pelatihan dan pengujian model. Namun, deteksi gerakan kompleks atau gerakan dengan variasi posisi tangan masih menjadi tantangan, meskipun hanya menyebabkan sedikit kesalahan klasifikasi

Saran

Untuk penelitian selanjutnya, disarankan beberapa langkah untuk meningkatkan efektivitas deteksi:

Variasi Data Pelatihan: Menambahkan variasi data pelatihan untuk mencakup gerakan yang lebih kompleks dan beragam.

Data Augmentation: Menggunakan metode data augmentation untuk memperkaya dataset yang tersedia.

Arsitektur Model Canggih: Mengeksplorasi arsitektur model yang lebih canggih untuk mengatasi tantangan klasifikasi gerakan sulit.

Optimasi Parameter Model: Mengoptimalkan parameter model untuk mempercepat pelatihan dan meningkatkan akurasi.

Dengan pengembangan lebih lanjut, sistem ini diharapkan menjadi solusi andal dan efektif dalam mendukung komunikasi berbasis bahasa isyarat.

DAFTAR PUSTAKA

- [1] I. I. Arifah, F. N. Fajri, and G. Q. O. Pratamasunu, "Deteksi Tangan Otomatis Pada Video Percakapan Bahasa Isyarat Indonesia Menggunakan Metode YOLO Dan CNN," *J. Appl. Inform. Comput.*, vol. 6, no. 2, pp. 171–176, Nov. 2022, doi: 10.30871/jaic.v6i2.4694.
- [2] World Health Organization, Addressing the rising prevalence of hearing loss. Geneva: World Health Organization, 2018. Accessed: Nov. 29, 2024. [Online]. Available: <https://iris.who.int/handle/10665/260336>
- [3] N. Adaloglou et al., "A Comprehensive Study on Deep Learning-Based Methods for Sign

- Language Recognition,” *IEEE Trans. Multimed.*, vol. 24, pp. 1750–1762, 2022, doi: 10.1109/TMM.2021.3070438.
- [4] A. Mittal, P. Kumar, P. P. Roy, R. Balasubramanian, and B. B. Chaudhuri, “A Modified LSTM Model for Continuous Sign Language Recognition Using Leap Motion,” *IEEE Sens. J.*, vol. 19, no. 16, pp. 7056–7063, Aug. 2019, doi: 10.1109/JSEN.2019.2909837.
- [5] C. K. M. Lee, K. K. H. Ng, C.-H. Chen, H. C. W. Lau, S. Y. Chung, and T. Tsoi, “American sign language recognition and training method with recurrent neural network,” *Expert Syst. Appl.*, vol. 167, p. 114403, Apr. 2021, doi: 10.1016/j.eswa.2020.114403.
- [6] B. Triwijoyo, L. Karnaen, and A. Adil, “Deep Learning Approach For Sign Language Recognition,” vol. 9, pp. 12–21, Mar. 2023, doi: 10.26555/jiteki.v9i1.25051.
- [7] A. Akdag and O. K. Baykan, “Enhancing Signer-Independent Recognition of Isolated Sign Language through Advanced Deep Learning Techniques and Feature Fusion,” *Electronics*, vol. 13, no. 7, p. 1188, Mar. 2024, doi: 10.3390/electronics13071188.
- [8] M. C. Bagaskoro, F. Prasajo, A. N. Handayani, E. Hitipeuw, A. P. Wibawa, and Y. W. Liang, “Hand image reading approach method to Indonesian Language Signing System (SIBI) using neural network and multi layer perseptron,” *Sci. Inf. Technol. Lett.*, vol. 4, no. 2, Art. no. 2, Nov. 2023, doi: 10.31763/sitech.v4i2.1362.
- [9] A. Duarte et al., “How2Sign: A Large-Scale Multimodal Dataset for Continuous American Sign Language,” presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 2735–2744. Accessed: Nov. 11, 2024. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2021/html/Duarte_How2Sign_A_Large-Scale_Multimodal_Dataset_for_Continuous_American_Sign_Language_CVPR_2021_paper.html?ref=https://githubhelp.com
- [10] K. Goyal, “INDIAN SIGN LANGUAGE RECOGNITION USING MEDIAPIPE HOLISTIC”.
- [11] G. H. Samaan et al., “Mediapipe’s Landmarks with RNN for Dynamic Sign Language Recognition,” *Electronics*, vol. 11, no. 19, p. 3228, Oct. 2022, doi: 10.3390/electronics11193228.
- [12] P. Selvaraj, G. NC, P. Kumar, and M. Khapra, “OpenHands: Making Sign Language Recognition Accessible with Pose-based Pretrained Models across Languages,” Oct. 12, 2021, arXiv: arXiv:2110.05877. doi: 10.48550/arXiv.2110.05877.
- [13] K. Roh, H. Lee, E. J. Hwang, S. Cho, and J. C. Park, “Preprocessing Mediapipe Keypoints with Keypoint Reconstruction and Anchors for Isolated Sign Language Recognition”.
- [14] R. M. Abdulhamied, M. M. Nasr, and S. N. A. Kader, “Real-time recognition of American sign language using long-short term memory neural network and hand detection,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 30, no. 1, Art. no. 1, Apr. 2023, doi: 10.11591/ijeecs.v30.i1.pp545-556.
- [15] P. Sheth and S. Rajora, Sign Language Recognition Application Using LSTM and GRU (RNN). 2023. doi: 10.13140/RG.2.2.18635.87846.
- [16] Y. Obi, K. Claudio, V. Budiman, S. Achmad, and A. Kurniawan, “Sign language recognition system for communicating to people with

- disabilities,” *Procedia Comput. Sci.*, vol. 216, pp. 13–20, Jan. 2023, doi: 10.1016/j.procs.2022.12.106.
- [17] R. S. L. Murali, “Sign Language Recognition System Using Convolutional Neural Network And ComputerVision,” vol. 4, no. 4, 2022.
- [18] D. Bragg et al., “Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective,” in *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*, in *ASSETS '19*. New York, NY, USA: Association for Computing Machinery, Oct. 2019, pp. 16–31. doi: 10.1145/3308561.3353774.
- [19] N. C. Camgoz, O. Koller, S. Hadfield, and R. Bowden, “Sign Language Transformers: Joint End-to-End Sign Language Recognition and Translation,” presented at the *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10023–10033. Accessed: Nov. 11, 2024. [Online]. Available: https://openaccess.thecvf.com/content_CVPR_2020/html/Camgoz_Sign_Language_Transformers_Joint_End-to-End_Sign_Language_Recognition_and_Translation_CVPR_2020_paper.html
- [20] S. Jiang, B. Sun, L. Wang, Y. Bai, K. Li, and Y. Fu, “Skeleton Aware Multi-Modal Sign Language Recognition,” presented at the *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 3413–3423. Accessed: Nov. 11, 2024. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2021W/ChaLearn/html/Jiang_Skeleton_Aware_Multi-Modal_Sign_Language_Recognition_CVPRW_2021_paper.html
- [21] H. Zhou, W. Zhou, Y. Zhou, and H. Li, “Spatial-Temporal Multi-Cue Network for Continuous Sign Language Recognition,” *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 07, Art. no. 07, Apr. 2020, doi: 10.1609/aaai.v34i07.7001.
- [22] M. Ismail, S. Dawwd, and F. Ali, “Static hand gesture recognition of Arabic sign language by using deep CNNs,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 24, p. 178, Oct. 2021, doi: 10.11591/ijeecs.v24.i1.pp178-188.
- [23] D. Li, C. Rodriguez, X. Yu, and H. Li, “Word-level Deep Sign Language Recognition from Video: A New Large-scale Dataset and Methods Comparison,” presented at the *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 1459–1469. Accessed: Nov. 11, 2024. [Online]. Available: https://openaccess.thecvf.com/content_WACV_2020/html/Li_Word-level_Deep_Sign_Language_Recognition_from_Video_A_New_Large-scale_WACV_2020_paper.html