Data in Brief 65 (2026) 112511

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

A dynamic Malaysian sign language dataset for sign language recognition and translation

Yuan Ting Chong[a], Yessane Shrrie Nagendhra Rao[a], Rehman Ullah Khan[a,*], Chee Siong Teh[a], Mohamad Hardyman Barawi[a], Mohd Shahrizal Sunar[b], Joan Jo Jo Sim[c]

[a] Faculty of Cognitive Sciences and Human Development, Universiti Malaysia Sarawak, Sarawak, Malaysia
[b] School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, Johor Bahru, Malaysia
[c] Faculty of Social Sciences and Humanities, Universiti Malaysia Sarawak, Sarawak, Malaysia

ARTICLE INFO

Article history:
Received 18 December 2025
Revised 19 January 2026
Accepted 20 January 2026
Available online 29 January 2026

Dataset link: BIM-SSD-V1 Dataset(Malaysian Sign Language) - DOI link (Original data)
Dataset link: BIM-SSD-V2 Dataset(Malaysian Sign Language) - DOI link (Original data)

Keywords:
Deaf and hard-of-hearing
Sign language dataset
MediaPipe keypoints
Gloss
Sign-to-gloss-to-text framework

ABSTRACT

Sign languages all around the world are unique and diverse. Each sign language shows the differences in cultural nuances of its origin locale giving it is distinctive nature. Thus, despite the positive outcomes of sign language recognition and translation research that has been widely conducted worldwide, there are still notable limitations to each system which are mainly caused by data limitations. The sign language recognition and translation research in Malaysia especially has been set back by the limited size and nature of datasets available that are concurrent with current technological developments. The current datasets available for Malaysian Sign Language (BIM – Bahasa Isyarat Malaysia) are small and limited to fingerspelling of alphanumeric characters and several dynamic words and short phrases. However, given the continuous nature of the sign language communication, these data are not enough to properly train machine learning models to recognize and translate continuous real-world signs. Therefore, in order to address this issue, we introduce a dynamic BIM dataset which comprises of video, gloss, and translation data consisting of alphanumeric characters, dynamic words and short phrases, and con-

* Corresponding author.
E-mail address: krullah@unimas.my (R.U. Khan).
Social media: @DrRehmanUllah (R.U. Khan)

https://doi.org/10.1016/j.dib.2026.112511

tinuous sentences. The dataset is split into two versions. The first version, BIM-SSD-V1 dataset comprises of 4,858 parallel video (RGB frames), gloss, and translation data while the second version, BIM-SSD-V2 dataset comprises of 3,143 parallel video (RGB frames), keypoints and gloss data for recognition purposes, and 4,900 parallel gloss and translation data for translation purposes. The raw videos are also available in the dataset. The dataset was developed and compiled with the help of the Deaf and Hard-of-Hearing community. This process also included the development of a Sign Language Module (translations for the video and gloss data) to assist in the development of the dataset. The image and video data were collected using smartphones and the respective gloss annotations for the data were prepared with the help of a BIM expert. The data collection process was designed to reflect everyday communication scenarios by incorporating varied sentence constructions, repeated signing instances, and recordings under different backgrounds and contextual conditions to introduce data-level variability relevant to real-world use. The total number of participants involved in the data collection process was four. There are also four samples for every character, word, phrase or sentence in the Sign Language Module. The dataset can mainly be reused by researchers who would like to conduct sign language recognition and translation research using the Sign-to-Gloss-to-Text framework. However, the dataset is not limited to only one framework and can be used for other sign language recognition and translation research frameworks accordingly.

Specifications Table

Subject	Computer Sciences
Specific subject area	Sign Language Recognition and Translation
Type of data	Raw and Processed .mp4 (Video Files), .png (Image Files), .csv (Comma Separated Values Files), .npy (NumPy Files), .txt (Text Files), .gloss (Gloss Files), .translation (Translation Files), .pdf (Portable Document Format File)
Data collection	The image and video data were captured in the controlled and uncontrolled environments with four signers. 4,858 samples were recorded using smartphone cameras. The signers were given a list of tasks which included the alphabet, numbers, words, and sentences. All data were labelled with their respective gloss annotations by a BIM expert. The videos were then split into sequential RGB frames and resized. These frames and their annotations files are stored as BIM-SSD-V1. A second version, BIM-SSD-V2 was created by merging the RGB frames for the alphabet, numbers, and words. There is a total of 3,143 RGB frame folders (with 2D-keypoints using MediaPipe and annotated glosses) and 4,900 glosses and their respective translations in BIM-SSD-V2.
Data source location	Sarawak Society for the Deaf, Lot 1862, Lorong Laksamana 8, Jalan Laksamana Cheng Ho, 93200 Kuching, Sarawak, Malaysia (Data Collected). Faculty of Cognitive Sciences and Human Development (FCSHD), Universiti Malaysia Sarawak, Jalan Datuk Mohammad Musa, 94300 Kota Samarahan, Sarawak, Malaysia (Data Stored).

(*continued on next page*)

Data accessibility	Repository name: Zenodo Data identification number: BIM-SSD-V1: 10.5281/zenodo.17573985 BIM-SSD-V2: https://doi.org/10.5281/zenodo.17587687 Direct URL to BIM-SSD-V1: https://doi.org/10.5281/zenodo.17573985 Direct URL to BIM-SSD-V2: https://doi.org/10.5281/zenodo.17587687 This dataset contains raw and processed data for sign language recognition and translation purposes and is only used for research. The data includes identifiable facial information of BIM signers. Therefore, it is not made available to the public. However, researcher(s) interested in using the BIM-SSD dataset for research purposes can make a request by completing the Data Usage Agreement [DUA] provided in the URL link below: https://drive.google.com/drive/folders/1feJJA3xBAL-bjB3NMzZ8oT40xHb8iFrC?usp=sharing Interested researcher(s) must submit the completed and signed DUA by contacting the authors via the email addresses provided in the DUA. The Researcher(s) must also provide supporting documentation (e.g., institutional ethics approval, research proposal, or supervisor endorsement) demonstrating that the requested use of the dataset is part of a legitimate research project. Submission of the DUA constitutes the researcher's agreement to comply with all terms and conditions outlined in the agreement. Responsibility for compliance with applicable ethical and legal requirements lies solely with the researcher(s).
Related research article	Y. S. N. Rao, Y. T. Chong, R. U. Khan, C. S. Teh, M. H. Barawi, S. Sunar, J. J. J. Sim, 2024. Dynamic sign language and translation through deep learning: A systematic literature review. J. Theor. Appl. Inf. Technol. 102, 21. https://www.jatit.org/volumes/Vol102No21/28Vol102No21.pdf

1. Value of the Data

- The dataset contributes two versions of dynamic Malaysian Sign Language (BIM – Bahasa Isyarat Malaysia) data for the development of dynamic sign language recognition and translation systems using deep learning models. It was prepared with the Deaf and Hard of Hearing (DHH) community to ensure systematic and comprehensive parallel video, gloss (labels or representations of the signs), and translation data were made available for researchers who are interested in developing sign language recognition and translation systems that can be used by the DHH community.
- It was carefully curated to include signers from different ethnics and signing styles to reflect the variability of individuals present in the real world. The BIM-SSD dataset provides static, isolated, and continuous data from a wide range of topics of daily communication in real-world scenarios with the emphasis being more on continuous data. This especially supports research in temporal modeling and sequence-to-sequence learning in deep learning models.
- This dataset can be used by researchers to build robust sign language recognition and translation systems for BIM. Besides, researchers can also use this dataset as benchmark dataset to train and test their models as the dataset was prepared in English. Despite the dataset being for BIM, the dataset was prepared in English to ensure researchers from all around the world can use this dataset in their studies so that the recognition and translation results can be validated by the researchers themselves without having to only depend on evaluation metrices.

2. Background

The BIM-SSD dataset was originally compiled to address the limited availability of dynamic sign language datasets for research, particularly for those that focus on signing as a continuous and temporal process rather than just static and isolated gestures. Due to the costly nature of curated dataset, most sign language-based systems research all around the world and in Malaysia

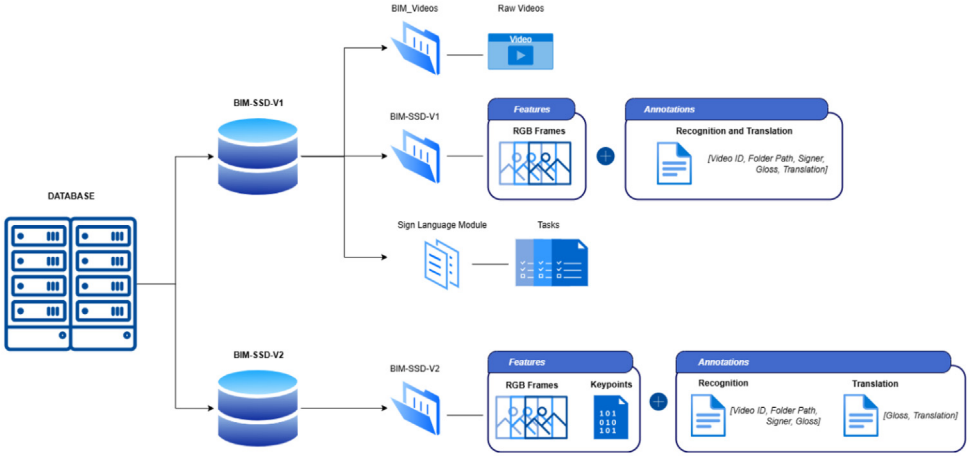

Fig. 1. BIM-SSD dataset data arrangement.

are restricted to using several benchmark datasets such as the RWTH-PHOENIX-Weather 2014T dataset [1] and the CSL-Daily dataset [2]. However, these datasets are limited to a specific sign language of a particular country, where the RWTH-PHOENIX-Weather 2014T dataset is based on the German language and the CSL-daily dataset is based on the Chinese language. The RWTH-PHOENIX-Weather 2014T dataset is also limited to a specific context which is regarding weather forecast [1].

Therefore, the BIM-SSD dataset was developed to serve as an additional resource for this field of research. The data collection was done based on the sign-to-gloss-text (S2G2T) framework to ensure that the dataset is suitable for processes that involve sign language gloss and end-to-end recognition and translation. This also ensures that the dataset is suitable for experimentation with emerging architectures and deep learning models for vision-based systems.

3. Data Description

Fig. 1 shows the overall data arrangement in the BIM-SSD dataset which is split into two versions, BIM-SSD-V1 and BIM-SSD-V2. These two versions can be found separately through their respective links as stated in the specifications table. In BIM-SSD-V1, there is a total of 4,858 images and video data that have been captured. The raw data can be found in the videos folder. The data were labelled as S1, S2, S3 and S4 with S1_A-1, S1_A-2, S2_B-1 and S3_D-10 being samples of image or video IDs. This means that there is a total of four different samples recorded for the same words, phrases or sentences by different signers with different backgrounds. The alphabet used in the labels represent different topics of real-world scenarios. For example, A refers to alphabets, B refers to numbers, while C refers to greeting phrases. Both versions of the dataset were split into train, dev and test sets, respectively.

Table 1 shows the content of the BIM-SSD-V1 and BIM-SSD-V2 dataset. The BIM-SSD-V1 dataset has RGB frames for all the 4,858 data samples. Each sequential frame was saved under their frame folders that were named with their video IDs. The BIM-SSD-V1 dataset was split into 4,458 RGB frame folders for train set and 200 RGB frame folders for dev and test sets, respectively. There are annotation files for each frame folder which have been labelled with their image and video IDs, frame folder name, signer ID, glosses, and translations. The BIM-SSD-V1 dataset was developed to support both sign language recognition and translation models for dynamic signs which include single words, phrases, and sentences. This is especially suitable for systems that process sign language images and videos based on the S2G2T

Table 1
Content of BIM-SSD-V1 and BIM-SSD-V2 datasets.

Dataset	Train	Dev	Test	Total	Process
BIM-SSD-V1	4,458	200	200	4,858	Recognition and Translation
BIM-SSD-V2	2,877	146	120	3,143	Recognition
	4,500	200	200	4,900	Translation

framework. However, this version of the dataset does not fully support continuous sign language recognition models as some frame folders only cover single isolated signs and their respective gloss.

The BIM-SSD-V2 dataset contains 3,143 RGB frame folders and was constructed specifically to support continuous sign language recognition (CSLR) models. Unlike BIM-SSD-V1, this version does not preserve the original recording structure. Instead, the isolated sign videos and their corresponding glosses were synthetically merged into groups to form longer sign sequences. For example, the isolated signs "SHOWER", "COMB", and "WATCH" were merged into a single sequential sample with the gloss sequence "SHOWER COMB WATCH". This design enables the training of CSLR models that require sequential inputs without relying on natural sentence-level recordings.

Importantly, BIM-SSD-V2 explicitly separates the recognition and translation data. The merged sequential samples are provided only for S2G recognition, which include parallel RGB frames, extracted 2D-keypoints and the corresponding gloss sequences. These recognition sequences were split into 2,877 train, 146 dev, and 120 test sets, respectively. Each frame folder was labelled with the respective video ID, frame folder name, signer ID, and gloss sequence in the annotation files.

For translation purposes, BIM-SSD-V2 provides a separate G2T data that contains 4,900 parallel gloss and translation pairs in lists. This component is intended exclusively for training G2T translation models and does not correspond directly to the merged video sequences. The translation data is divided into 4,500 lines for the training set, and 200 lines for both the dev and test sets, respectively. The initial lines in these files were 4870 lines which includes twelve lines for videos that were previously removed from the dataset due to errors. These sentences were kept for additional translation training purposes however, these do not affect any part of the data for the recognition aspect. Additionally, 30 more lines of parallel gloss and translation mainly for fingerspelling of common Malaysian names and signs of Malaysian honorifics were added to the training set to ensure the translation model can correctly translate names and honorifics. None of the names added refer to the participants in the dataset.

In typical SLRT pipelines, recognition and translation models are trained independently. Therefore, BIM-SSD-V2 is designed to support modular SLRT models' training within the S2G and G2T frameworks. Although the recognition and translation data in this version are not parallel, the dataset still remains suitable for training CSLR models and subsequent gloss-to-text translation models within staged or modular learning frameworks.

4. Experimental Design, Materials and Methods

The BIM-SSD dataset was developed by our team, consisting of two master students who are studying in the field of Applied Computing, their supervisor from the Faculty of Cognitive Sciences and Human Development (FCSHD), Universiti Malaysia Sarawak (UNIMAS), Malaysia, and a team from the Sarawak deaf society. The researchers cooperated with a group of BIM experts from the Sarawak Society for the Deaf (SSD) in developing the BIM-SSD dataset. SSD is a well-established organization representing the DHH community in Kuching, Sarawak. In order to ensure good communication between the researchers and the SSD team, a BIM signer who is also a hearing person was invited to be a mediator between them. Everyone involved in this

Table 2
Participant inclusion and exclusion criteria.

Criteria	Inclusion	Exclusion
Nationality	Malaysian	Non-Malaysian
Language Skill	Knowledgeable in BIM and able to communicate using BIM	No knowledge of BIM or unable to communicate using BIM
Age Range	Above 18 years old	Below 18 years old
Health and Physical Ability	Able to physically perform hand gestures and facial expressions	Unable to physically perform hand gestures and facial expressions
Ethnicity	Diverse ethnic groups (e.g. Malay, Chinese, Indian, Iban etc) which naturally reflects variations in skin tones and facial features	None
Appearance	Wear different types of attire or attires commonly to daily life usage	Strict adherence to any particular dress code
Facial Visibility and Comfort	Comfortable with showing their faces during recording	Uncomfortable with their faces being recorded
Availability	Must be able to attend the data collection session physically	Unable to attend scheduled data collection sessions

Table 3
Details of participants selected for the data collection process .

Signer ID	Ethnic	Age
PG1	Bidayuh	35
PG2	Chinese	32
PB1	Chinese	28
PB2	Malay	30

research project attended discussions and meetings before conducting the data collection process to develop the BIM-SSD dataset.

After discussion and confirmation, the researchers focused on selecting the BIM signers and preparing a Sign Language Module. Malaysia is a multicultural country, which has diverse ethnic groups such as Malays, Chinese, Indians and numerous indigenous groups. Thus, the researchers have listed the inclusion and exclusion criteria for selecting suitable BIM signers from the SSD team in Table 2 below. Meanwhile, Table 3 shows the four signers who aligned with the requirements listed and were selected as the participants for the data collection process.

The development of the BIM-SSD dataset involved three processes, which are task determination, data collection, and data organization. The data collection process was segregated through ten tasks with each task having two or more topics of static, isolated, and continuous data. For example, task one covered the alphabet, numbers, and greetings and basic phrases which are considered as three topics. All the ten tasks were compiled and documented in a Sign Language Module that was formulated for the development of this dataset. After the BIM signers were chosen, all of them gathered for discussions to check and validate the words, phrases and sentences that were listed in their tasks. The overall topics of the module are as shown in Table 4 below. The number of classes for each topic is stated in the table as well. Classes here refer to the number of words, phrases, and sentences that were prepared for each topic.

Each recording session involved the recording of one task. Hence, there were a total of ten sessions conducted. Once the tasks were prepared and validated, the data collection process was conducted. Image and video data were collected with the participants. The data were captured in controlled environments with white or plain backgrounds and uncontrolled environments with windows or outdoor settings as the backgrounds. There was a total of four signers who participated in the data collection process. These videos were captured with smartphone cameras which were the VIVO Y78 5G PLUS mobile phone with a rear camera of 68 megapixels with Optical Image Stabilization and a HUAWEI Y9 Prime 2019 mobile phone. Both mobile phones were positioned using a tripod stand and capture participants from front view. The BIM

Table 4
Topics and number of classes for each topic in the sign language module.

Topic	Number of Classes
Alphabet	26
Numbers	112
Greetings and Basic Phrases	17
Time, Days, Dates and Months of the Year	25
Everyday Communication and Social Interactions	56
Personal Information and Identification	26
Common Daily Routines	28
Weather	8
Health and Healthcare	16
Colours	20
Animals	50
Things	172
Feelings and Emotions	40
Hobbies and Activities	127
Occupations	62
Subjects	41
Celebrations and Wishes	18
Precaution, Safety, and Emergency Response	51
Basic Interview Conversations	19
States, Capitals and Places	79
Food and Dining	42
Directions, Transportation and Road Situations	43
Clothing and Shopping	43
Hotel and Travel	21
Bank	16
Services and Maintenance	6
Social Media	24
Gym	8
Public Etiquette and Behaviour	19

participants were showing their faces and hands clearly in the videos. A BIM expert from SSD also assisted the participants during the data collection sessions to ensure that signs signed by the participants are accurate and adhere to BIM conventions.

After capturing the data, all the data were then rechecked, verified and labelled with gloss annotations by the BIM expert. All the data and their gloss annotations were then made parallel to their respective translations in the Sign Language Module. Then, the researchers split the data into 4,458 videos for the training set and 200 videos equally for the dev and test sets. The initial number of samples collected was 4,860. However, two videos were removed from the dataset due to signing error. Therefore, the total number of samples in the dataset is 4,858. Fig. 2 below shows some gloss annotations of some of the videos and their translations from the training set.

Next, the researchers proceeded to preprocess the data. Firstly, all the videos were split into sequential RGB frames using OpenCV. The extracted frames were passed through the YOLO (You Only Look Once) algorithm. YOLO is a tool that is mostly applied in computer vision for object detection, recognition, and classification in real-time [3]. In this study, YOLOv8n with pretrained weights (yolov8n.pt) was selected to detect the BIM signers and crop them within the center of each frame.

For each frame, the bounding box was drawn around the detected signer and expanded by a fixed padding ratio of 15% in both width and height to ensure the hand and upper-body movements were not truncated. The cropped frames were continuously resized to a fixed spatial resolution of 256×256 pixels while preserving the original frame rate of 25 frame per second (fps). These RGB frames and annotation files can be found in the BIM-SSD-V1 dataset. Fig. 3 below shows an example of a RGB frame before and after preprocessing.

Using these pre-processed RGB frames as the base, the BIM-SSD-V2 dataset was subsequently constructed. After pre-processing, the researchers rechecked if all gloss annotations were paral-

```
name|video|start|end|speaker|orth|translation
S4_D5-16|S4_D5-16/*.png|-1|-1|PG2B2|DIRECTION POINT-ME|Can you point me in the right direction? .
S4_D5-17|S4_D5-17/*.png|-1|-1|PG2B2|EMERGENCY EXIT SHOW-ME|Can you show me the emergency exit? .
S4_D5-18|S4_D5-18/*.png|-1|-1|PG2B2|#QR PAYMENT YOU(PRO-2) ACCEPT|Do you accept QR payment? .
S4_D5-19|S4_D5-19/*.png|-1|-1|PG2B2|#QR YOU(PRO-2) HAS|Do you have QR? .
S4_D5-20|S4_D5-20/*.png|-1|-1|PG2B2|CARD USE CAN|Can I use my card? .
S4_D5-21|S4_D5-21/*.png|-1|-1|PG2B2|TNG PAY CAN|Can I pay with TnG? .
S4_D5-22|S4_D5-22/*.png|-1|-1|PG2B2|MONEY TRANSFER BACK|Can I transfer the amount to you instead? .
S4_D5-23|S4_D5-23/*.png|-1|-1|PG2B2|CASH LOW|I am low in cash .
S4_D5-24|S4_D5-24/*.png|-1|-1|PG2B2|#QR HAS|Can I have the QR? .
S4_S3-1|S4_S3-1/*.png|-1|-1|PG2B2|FOOD MENU HAS|Can I have the menu? .
S4_S3-2|S4_S3-2/*.png|-1|-1|PG2B2|FOOD MENU WITH PICTURE HAS|Do you have a menu with pictures? .
S4_S3-3|S4_S3-3/*.png|-1|-1|PG2B2|MONEY COUNT SHOW-ME|Can you show me the price on the calculator? .
S4_S3-4|S4_S3-4/*.png|-1|-1|PG2B2|#RECEIPT NEED|I need the receipt .
S4_S3-5|S4_S3-5/*.png|-1|-1|PG2B2|#RECEIPT GET PLEASE|Can I get the receipt, please? .
S4_S3-6|S4_S3-6/*.png|-1|-1|PG2B2|SHOW-ME DRINK WHERE|Can you show me where the drinks section is? .
S4_S3-7|S4_S3-7/*.png|-1|-1|PG2B2|WATER PLAIN SOME|I would like some plain water .
S4_S3-8|S4_S3-8/*.png|-1|-1|PG2B2|LAKSA ORDER|I would like to order laksa .
S4_S3-9|S4_S3-9/*.png|-1|-1|PG2B2|ICE SOME|Can I get some ice? .
```

Fig. 2. Gloss annotations and translation for videos from dataset.

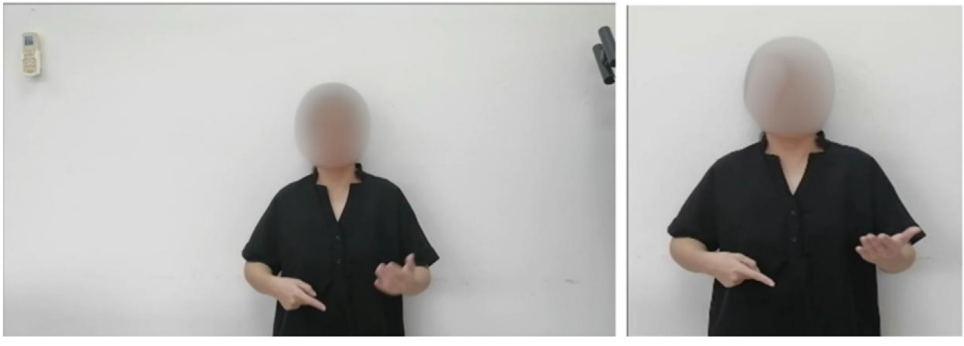

Fig. 3. RGB frame before and after preprocessing.

lel to the RGB frame folders. A total of 4,858 RGB frame folders were reviewed and divided into frame folders that included single and multiple glosses. The researchers rearranged all the single glosses by merging them into multiple gloss sequences. Then, the frame folders' names were updated based on the video IDs merged. For example, S1_A-1, S1_A-2, S1_A-3, and S1_A-4 that were initially single glosses "#A", "#B", "#C", and "#D" were merged into "#A #B #C #D" and renamed into video ID S1_A-1_A-2_A-3_A-4. This merging process of single glosses was important for the BIM-SSD-V2 dataset because this dataset was developed to support continuous sign language recognition that needs continuous signs being fed to the model. Only single glosses were converted into multiple gloss sequences through the merging process. The remaining RGB frames which are originally sequential glosses were kept as they are in this dataset.

This resulted in 3,143 RGB frames folders in the BIM-SSD-V2 dataset. These RGB frames were continuously passed for external features and keypoints extraction through MediaPipe. MediaPipe is a common tool that is used to extract keypoints for detecting human movements in this field [4,5]. The MediaPipe tool processed each frame independently and detected the human body, hands, and pose landmarks. MediaPipe was configured with a minimum of 0.5 confidence detection and operated without temporal smoothing across frames.

For each frame, these landmarks were then converted into numerical coordinate, which were 33 body pose keypoints and 21 keypoints for each hand. Each extracted body pose resulted in 132 pose and 63 per hand features through (x,y,z) coordinates and a visibility score rep-

Fig. 4. Extracted keypoints in RGB frames format.

resentation. When pose or hand landmarks were not detected in a frame, the corresponding keypoint values were filled with zeros to maintain a fixed feature dimension. The outputs for the keypoints were shaped into [number of frames (T), 258] in a NPY file for each RGB frame folder. Fig. 4 below shows examples of keypoints extracted from a signer in the RGB frames format.

Once the keypoints' features were prepared, the researchers rearranged the gloss annotation files and split the frame folders into train, dev, and test sets again. The annotation files for this dataset covered the recognition part from signs to glosses for each frame folders. The translation part was prepared separately for the gloss sequences with their respective translation according to the lines in the gloss list. These parts are separated because there is no valid translation that can be mapped to the merged single glosses as there would be grammatical error. Therefore, both recognition and translation in the BIM-SSD-V2 dataset are divided into different sections. For recognition, there are 2,877 frame folders for train set, 146 frame folders for dev set, and 120 frame folders for test sets. However, there are 4,500 lines for train set, and 200 lines for dev and test sets, respectively which cover the entire list of glosses and their respective translations. Researchers may use either the BIM-SSD-V1 or BIM-SSD-V2 dataset according to the requirements of their research.

Limitations

Number of Participants: There was only a total of four signers who participated in the dataset development. Two of them are male and two of them are female. This may cause signer bias as the generalization ability of sign language recognition and translation models become limited.

Camera Viewpoint: All the videos recorded in this dataset only cover a single front-facing camera angle. This means that there may be hand overlaps and side facial expressions that were not clearly captured which may cover some external semantic or grammatical information.

Gloss Level Representation: The BIM-SSD-V1 dataset contained single words and short phrases which needed to be modified through merging. Since these words and phrases were originally annotated as single glosses, this caused limitation in supporting CSLR. This limitation was solved through the development of BIM-SSD-V2 dataset where the recognition data was synthetically concatenated. This has caused a lack of natural language alignment of the sentences and the merged glosses also have no valid translation information.

Dataset Scale: This dataset was novelly built for BIM but it still relatively small when compared to large-scale sign language datasets, such as the benchmark RWTH-PHOENIX Weather 2014T (Camgoz et al., 2018) and CSL-Daily (Zhou et al., 2021) datasets. This may restrict the learning of larger deep learning models due to a smaller number of samples.

Ethics Statement

Informed consent was given and obtained from all participants involved in the preparation of this dataset. Apart from the participants' faces and upper bodies recorded in the dataset, no other personal information was obtained or made publicly available. This research was approved by the Universiti Malaysia Sarawak Human Research Ethics Committee (Non-Medical), HREC(NM)/2023 (2)/33 and HREC(NM)/2023 (2)/49.

CRediT Author Statement

Yuan Ting Chong: Conceptualization, Methodology, Software, Investigation, Data curation, Writing – Original Draft; **Yessane Shrrie Nagendhra Rao:** Conceptualization, Methodology, Software, Investigation, Data curation, Writing – Original Draft; **Rehman Ullah Khan:** Conceptualization, Validation, Supervision, Writing – Review and Editing; **Chee Siong Teh:** Supervision; **Mohamad Hardyman Barawi:** Supervision; **Mohd Shahrizal Sunar**: Supervision; **Joan Jo Jo Sim:** Resources, Supervision.

Data Availability

BIM-SSD-V1 Dataset(Malaysian Sign Language) - DOI link (Original data) (Zenodo)
BIM-SSD-V2 Dataset(Malaysian Sign Language) - DOI link (Original data) (Zenodo)

Acknowledgments

We gratefully acknowledge the support and resources provided by the Research, Innovation and Enterprise Centre (RIEC) and the Faculty of Cognitive Sciences and Human Development, Universiti Malaysia Sarawak. We would also like to acknowledge the Sarawak Society for the Deaf (SSD) for their support and contribution to this research.

Funding Statement

This research received a Fundamental Research Grant (FRGS) NAT/F04/FRGS/86033/2023 from the Malaysian Ministry of Higher Education.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] N.C. Camgoz, S. Hadfield, O. Koller, H. Ney, R. Bowden, Neural sign language translation, in: Proc. 2018 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 7784–7793, doi:10.1109/cvpr.2018.00812.
[2] H. Zhou, W. Zhou, W. Qi, J. Pu, H. Li, Improving sign language translation with monolingual data by sign back-translation, in: Proc. 2021 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2021, pp. 1316–1325, doi:10.1109/cvpr46437.2021.00137.
[3] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: unified, real-time object detection, ArXiv (2016) 1–10 https://arxiv.org/abs/1506.02640.

[4] T. Adão, J. Oliveira, S. Shahrabadi, H. Jesus, M. Fernandes, A. Costa, V. Ferreira, M.F. Goncalvas, M.A.G. Lopez, E. Peres, L.G. Magalhaes, Empowering deaf-hearing communication: exploring synergies between predictive and generative AI-based strategies towards (Portugese) sign language interpretation, J. Imaging 9 (2023) 11, doi:10.3390/jimaging9110235.
[5] J.M. Joshi, D.U. Patel, Dynamic Indian sign language recognition based on enhanced LSTM with custom attention mechanism, SSRG Int. J. Electron. Commun. Eng. 12 (2024) 2, doi:10.14445/23488549/IJECE-V11I2P107.