INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

OPEN ACCESS

Received: 15-07-2024

Accepted: 01-11-2024

Published: 30-11-2024

Citation: Muthusamy P, Murugan GP (2024) Occlusion Resistant Spatio-Temporal Hybrid Cue Network for Indian Sign Language Recognition and Translation. Indian Journal of Science and Technology 17(44): 4590-4599. https://doi.org/10.17485/IJST/v17i44.2225

* **Corresponding author**.

mprema.hanuphd@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Occlusion Resistant Spatio-Temporal Hybrid Cue Network for Indian Sign Language Recognition and Translation

Prema Muthusamy[1*], **Gomathi Pudupalayam Murugan**[2]

1 Research Scholar, Department of Computer Science, P.K.R Arts College for Women, Erode, 638476, Tamil Nadu, India
2 Associate Professor and Head, Department of Computer Science, P.K.R Arts College for Women, Erode, 638476, Tamil Nadu, India

Abstract

Objective: To tackle the issues of occlusion in human skeleton extraction and simplify the pixel matching related to the human skeleton structure for efficient Indian Sign Language (ISL) recognition and translation. **Methods:** This paper presents Occlusion-Resistant STHCN (OSTHCN) to tackle the occlusion problem in human skeleton extraction for effective ISL recognition and translation. This model incorporates Skeleton Occupancy Likelihood Map estimation using B-Spline curves to enhance the skeleton extraction. Due to occlusions caused by fingers and hands, the extracted skeleton is composed of disconnected skeletal subgraphs. Consequently, each observed skeleton is represented as a probability distribution along an ellipsoidal contour, originating from the central points of the skeleton. A heuristic technique estimates occluded skeletons using 3D probability map with an occupancy grid where each voxel indicates skeleton likelihood. The occupancy distribution is updated using observed branch clusters across image sequences, detecting occluded skeletons by finding minimum-cost paths. Finally, Maximally connected subgraphs are merged into a main graph by finding minimum-cost paths in the 3D likelihood map, enabling the prediction of occluded skeleton parts for ISL recognition and translation. **Findings:** OSTHCN model achieved an accuracy of 96.74% on the ISL Continuous Sign Language Translation Recognition (ISL-CSLTR) dataset outperforming existing prediction models. **Novelty:** This model employs a unique occlusion-handling strategy for skeleton extraction, estimating occluded part, integrating connected subgraphs via minimal cost path searches for more precise skeleton parts and enhancing accuracy for ISL recognition and translation.

Keywords: Sign Language; Graph Convolutional Neural Network; Ellipsoidal Contour; 3D Likelihood Map; Occupancy Probability

1 Introduction

Sign Language Recognition (SLR) aims to understand the gestures of deaf-mute individuals by identifying features and classifying signs as gestures [1]. The primary goal of SLR is to translate these signs into spoken or written language without manual interpretation. Continuous Sign Language Recognition (CSLR) is a key aspect of SLR, as it captures the contextual relationships between signs which significantly impacts the meaning of sentences and reduces the need to identify individual gestures [2]. CSLR datasets typically consist of videos featuring sign sequences rather than isolated signs, making them better suited for practical applications and enhancing the overall effectiveness of SLR [3]. Recognizing CSLR is complex due to variations in speed context and overlapping gestures further complicating accurate interpretation.

Recently, Deep Learning (DL), a subset of Artificial Intelligence (AI) has emerged as a powerful tool for improving SLR. DL models including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) and Deep Belief Networks (DBN) effectively analyze the data to identify sign language phrases and translate them to text, facilitating better communication between signers and non-signers [4]. DL models are a promising option for advancing CSLR applications, efficiently understanding sign communication semantics and improving sign language prediction for individuals with hearing impairments [5]. Some of the DL based SLR models are provided below.

Sharma & Singh [6] devised a spatio-temporal framework for Dynamic ISL recognition using a Stacked 2 Bidirectional LSTM (S2B-LSTM) model. CNN extracts spatial features from input video data, while S2B-LSTM captures temporal relations between consecutive frames. The model was trained and tested on a self-developed dataset of ISL dynamic gesture videos. Muthusamy & Murugan [7] developed Spatio-Temporal Hybrid Cue Network (STHCN) using Dynamic Dense Spatio-Temporal Graph CNN (DDSTGCNN) and VGG11+1D-CNN+BLSTM for ISL recognition and translations. The skeleton data is obtained from input videos and inputted to DDSTGCNN which learns spatial features. VGG11+1D-CNN+BLSTM retrieves features from full frame input and inputs them into BLSTM encoders, CTC decoders and SA-LSTM decoders for sequence learning and inference, enhancing the prediction and translation of ISL. Cui et al. [8] developed an end-to-end CSLR network, the Spatial–Temporal Transformer Network (STTN) to process sign language videos. By chunking video frames into patches, this model reduces computational complexity. The global, spatial action and semantic temporal features were retrieved using visual feature extraction. Cross-entropy loss was used to align the video and text sequences.

Sharma et al. [9] developed a novel DL model for ISL recognition. In this method, the extracted sign language actions were converted into the text of users. Then, VGG16 was employed for sign language recognition. Natural Language Processing (NLP) algorithms were utilized to translate text input from users into sign language. Vidhyalakshmi et al. [10] developed an ISL model using EfficientNetB0 (ISLR-EfficientNetB0) with transfer learning. Input images from the ISL dataset undergo pre-processing using the Fairness-Aware Collaborative Filtering (FACF) which removes noise and resizes images. Finally, Efficient Net B0 was employed to identify ISL. Priya & Sandesh [11] suggested a DL model for offline and real-time ISL recognition system. This model aids communication for the speech and hearing impaired by recognizing ISL digits and alphabets from two datasets, with and without preprocessing. Machine Learning (ML) was used for Hue, Saturation and Value (HSV) conversion, skin mask generation and Gabor filtering for segmentation. Finally, YOLO-NAS-S was employed for real-time ISL recognition.

1.1 Research Gap

The previously mentioned models do not sufficiently address the challenges that occlusions present in the human body pose, making it difficult for certain pixels to align with the skeleton structure during detection. This complicates the extraction of clear human skeleton points, as occlusions further hinder the matching of pixels to the human skeleton structure within the network model, ultimately impairing effective detection and extraction. In this context, this study aims to identify the missing skeleton points. It manages disjoint subgraphs to account for variations in hand and finger positioning, thereby enhancing performance in ISL recognition and translation. This model provides a more advanced method for handling occlusions compared to traditional approaches.

1.2 Major Contributions

This article proposes OSTHCN for effective ISL recognition and translation. This model incorporates a Skeleton Occupancy Likelihood Map estimation using B-Spline curves to enhance the skeleton extraction process of DDSTGCNN [7]. Because of the occlusion by fingers and hands, the retrieved skeleton is made up of unconnected skeletal subgraphs of a single skeleton. Individually, the observed skeleton is identified as probability distributions of an ellipsoidal contour beginning at the skeleton point of centers. The model employs heuristic assumptions to predict the likelihood of occluded skeletons by creating a 3D occupancy grid probability map. The occupancy probability distribution of a 3D map is changed using learned branch clusters over various image sequences, recognizing occluded skeletal components through low-cost paths. By looking for minimal cost pathways in the 3D likelihood map to anticipate the occluded regions of the skeleton for ISL identification and translation, the maximally connected subgraphs are merged into the main graph. The maximum connected subgraphs are combined into the main graph by looking for the lowest cost pathways in the 3D probability map to forecast the occluded sections of the skeleton for ISL identification and translation.

2 Methodology

In this section, the proposed OSTHCN model is briefly illustrated. Figure 1 depicts the block structure of the proposed model.

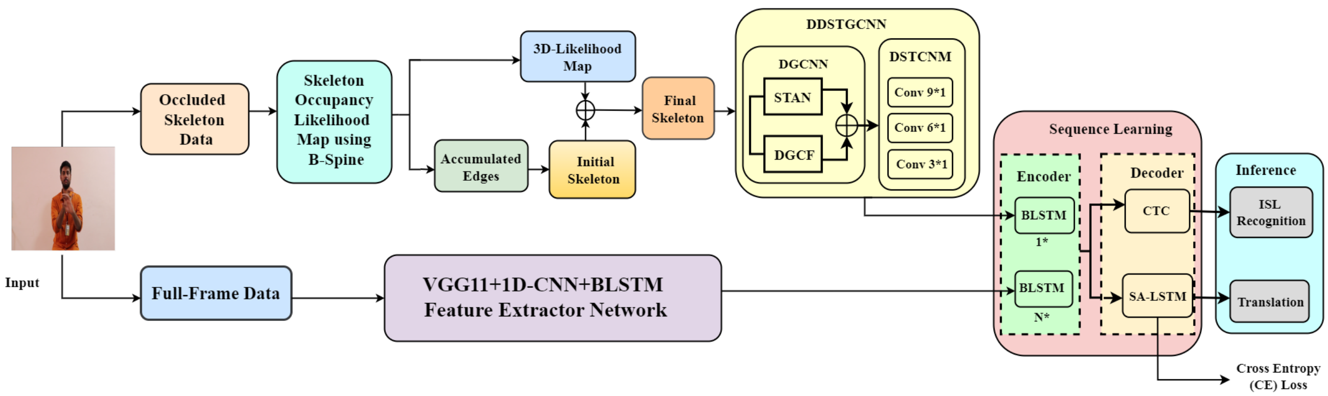

Fig 1. Pipeline of the Proposed Model

2.1 Dataset Description

For experimental purposes, ISL-CSLTR[12] is employed in this paper. The ISL-CSLTR corpus comprises a large dataset of 700 fully annotated videos, including 18,863 sentence-level frames and 1,036 word-level images, representing 100 spoken language sentences performed by 7 different signers. It is organized by signer variations and time boundaries, with complete annotations made publicly available. The main goal of this corpus is to facilitate research in Sign Language Translation and Recognition (SLTR), providing a comprehensive resource for building frameworks to convert spoken language into sign language and vice versa. By addressing key challenges in SLTR, this corpus enhances translation and recognition capabilities, offering clear annotations that aid researchers in understanding the data.

2.2 Extraction of Semantic Information and Key Annotations

Semantic information and key annotations are extracted from the ISL-CSLTR dataset. A detailed explanation of the extraction process for efficient ISL recognition is provided in[7].

2.3 Semantic 3D Points

The model intends to provide better comprehension of occluded human body areas, which are often self-occluded by hands or fingers. It creates a 3D probability map in the form of an occupancy grid, gathering information about visible and occluded objects in order to extract skeleton images. Observed branch clusters over pictures are used to alter the occupancy probability distribution. The final skeleton is built by joining previously grouped branches in the probability map using lowest cost route searches.

2.4 Skeleton Occupancy Likelihood Map

A 3D point is obtained semantically to mark visible sections of the human body such as the face, hands, and fingers in which the branches are not occluded. This network is trained using annotated images and actual and synthetic images. Each branch instance is labeled as a slender polygon, allowing line-fitting on point clusters. A new technique is presented for estimating the existence of occluded branches in masked skeletons using 3D likelihood map in the form of an occupancy grid, where the x^{th} grid voxel maintains the skeleton occupancy probability of $P_L\left(v_x\right) \in (0;1]$. Each branch pattern is projected to extend in a given apparent branch instance along with its development orientation using the notion that branches flow on realistic human movements. This estimate is based on a heuristic assumption. It is characterized as a branch point cloud cluster of independently observed probability variations P_0 with an ellipsoidal contour.

In Figure 2(**A**), the human skeleton's segmented branch component with a 0.89 confidence level is represented by three-dimensional set of points. The principal components (C_x) are identified by centering a B-spline curve with four control points (CP_x) as in Figure 2(**B**). After obtaining P_0 at each segment of the line in Figure 2(**C**), P_L is then updated in Figure 2(**D**) by using Equation (2). Line segments are recovered from clusters of identified branches throughout the images and the resulting joint 3D likelihood map is shown in Figure 3.

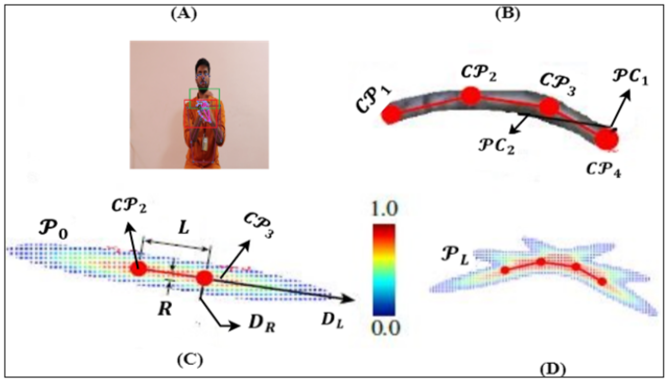

Fig 2. Skeleton Occupancy Likelihood Map using B-Spline Curve

The least squares approach is used to estimate a 3D-B-Spline curve on a point cloud using four control points. This process accumulates over images to form the skeletal structure of the point cloud cluster depicted in Figure 2(**B**), represented by three connected line segments which is depicted in Figure 3. The diameter of the branch instance is estimated using principle component analysis (PCA), joint likelihood map's probability (P_0) distribution is improvised P_0 by according to predefined criteria.

- The P_0 of the voxels containing the line segment serves as a representation of the mask confidence score from the instance segmentation network.
- An ovoid contour and line partition with decreasing axial and peripheral dimensions are next to the voxel P_0 in Figure 2(C), which is shown in Equation (1).

$$P_0 = C - \frac{C}{T}\sqrt{\left(\frac{D_L}{L}\right)^2 + \left(\frac{D_R}{R}\right)^2} \tag{1}$$

Where,

C – Mask Confidence Outcome,

L – Line Segment Length,

R – Approximate Radius

D_L & D_R – Voxel's axial and radial distance from the line segment

T – Used to ontrols the rate at which P_0 declines when D_L and D_R increases.

$$P_L\left(v_x\right) = 1 - \left[1 - P_L\left(v_x\right)\right]\left[1 - P_0\left(v_x\right)\right] \tag{2}$$

In the 3D likelihood map, P_0 from all lines increases the occupancy possibility in voxels immediately traversed by a line segment and in voxels concurrently impacted by many lines.

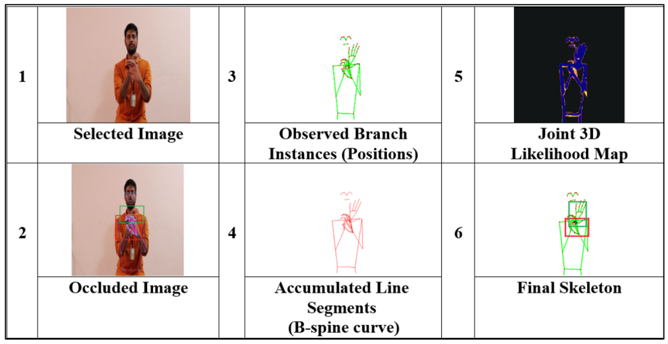

Fig 3. Overall 3D Joint Likelihood Map Representation

2.5 Extracting the skeleton views

An eventual tree skeleton $G_Z = (V_Z, E_Z)$ is an unstructured circular graph with vertices V_Z and edges E_Z that represents the hierarchical layout of the skeleton canopy. The joint likelihood map is utilized to post-process the accumulated line segments, encapsulating the gesture relation of the branch clusters. The first tree skeleton is generated as $G_f = \left(V_f, E_f\right)$ by aggregating the accumulating line segments. The Laplacian smoothing technique is a method used to transform a set of points into a skeleton curve by sampling points evenly along a line and adjusting spacing to fine-tune the voxel size of the probability map is illustrated as follows,

$$J_x = \frac{1}{Q} \sum_{y=1}^{m} J_y \tag{3}$$

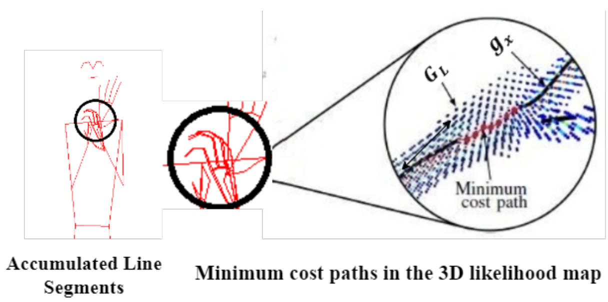

Fig 4. Identifying Cost paths in Likelihood Map

This new location of the vertices with respect to x is denoted by J_x, where Q is the number of neighboring vertices, J_y is the position of neighbor vertex y and so on. Applying Laplacian smoothing iteratively until the sum of the changes in the vertex position of each iteration converges below the threshold value produces an ensemble of improved skeletal edges. To make E_f, we start with V_f and generate a euclidean minimum spanning tree with edges greater than the voxel size along each vertex. Due to hand and finger occlusion, the graph $G_f = \{g_x\}$ is composed of independent skeletal subgraphs g_x of a single tree. Figure 4 shows an example of lowest cost-route in 3D probability map.

Figure 4 depicts the process of integrating maximally connected subgraphs in G_f by searching for the cheapest routes in a 3D probability map in order to predict the hidden skeleton components. Initial steps include transforming the probability map into a weighted graph G_L by including all of the undirected edges between neighboring nodes. Also, Figure 4 depicts the primary skeleton G_f as black-colored curves overlapping with the weighted probability graph G_L. Due to occlusion, G_f is made up of discontinuous subgraphs g_x. The lowest cost pathways have been explored in g_x to coordinate with the unconnected subgraphs g_x. According to Equation (4), the inverse log of the mean occupancy probability is used to adjust the cost of an edge connecting vertices a and b.

$$C\left(e_{ab}\right) = \frac{-log(P_L\ (v_a) + P_L\ (v_v))}{2} \tag{4}$$

As a result of Equation (4), the cost of edges between highly occupied nodes is minimized. The cheapest path among all possible permutations of paired subgraphs in G_Z is found using the weighted probability graph G_L, which is initialized with G_f. The cheapest route is supplemented by G_Z. The final skeleton output G_Z is comprised of both the occlusion-obscured skeletal curves that are immediately apparent and the inferred skeletal curves that are not immediately apparent, and this procedure is repeated until no more pathways are detected. The algorithm constructed for the likelihood map path search is depicted below.

Algorithm: Skeleton Occupancy Likelihood Map Estimation

Input: Eventual skeleton G_f, Likelihood Map P_L

Output: Human skeleton G_Z

1. $G_Z \leftarrow G_f$
2. $G_L \leftarrow WeightundirectedGraph\,(P_L)$
3. Repeat
4. $s \leftarrow EmptyArray\,()$
5. Complete pairwise subgraph integration $(g_a, g_b) \in G_Z$ do
6. $\rho \leftarrow DijkstraMinPath\,(g_a, g_b,\, G_Z)\, do$
7. P. Append (ρ)
8. $end\ for$
9. $\rho_{min} \leftarrow Mincostpath\,(\,P)$
10. $G_Z \leftarrow Joinsubgraphs\,(G_Z,\, \rho_{min})$
11. Until no more paths are determined

Hence, by utilizing the OSTHCN model, a clear human skeleton point model is retrieved and reduces the occlusion issues during human skeleton extraction for the recognition and translation of ISL. Figure 5 represents the flowchart of this methodology.

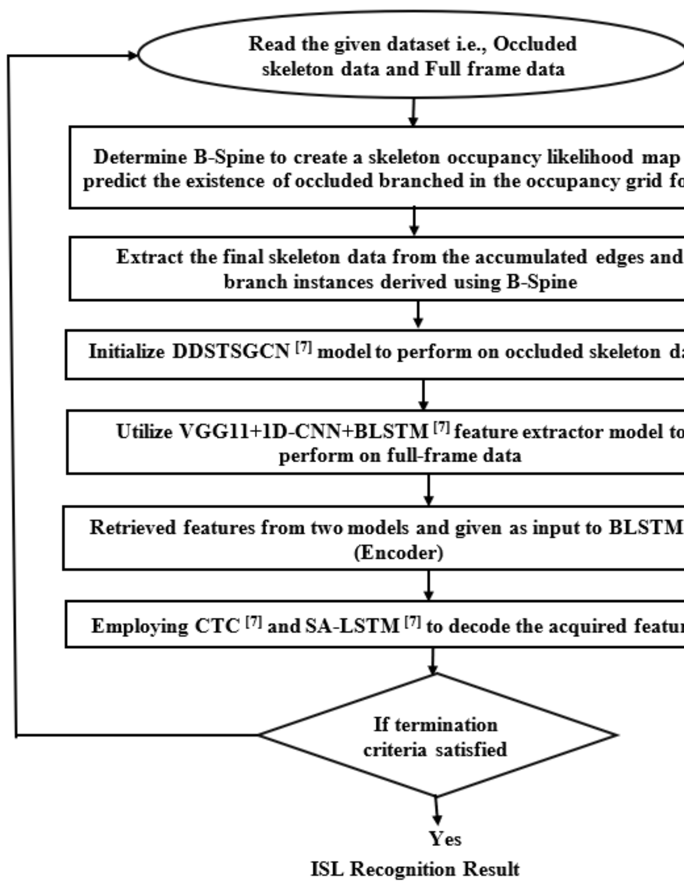

Fig 5. Flowchart for the proposed model

3 Results and Discussion

In this section, the efficiency of the OSTHCN model is examined by implementing it in Python using the dataset which is discussed in Section 4.1. For experimental purposes, 610 videos from the ISL-CSLTR[12] dataset have been gathered, with 60% (366) of the data used for training and the 40% (244) used for testing. From the collected dataset 86 sentences (labels) and 25 per second time frame rate have been determined for the final output. Further, a comparative analysis is carried out to understand the improvement of the OSTHCN, model contrasted to the existing models including S2B-LSTM[6], STHCN[7], STTN[8], VGG16[9],ISLR-EfficientNetB0[10] and YOLO-NAS-S[11]. The assessment measures that are utilized to evaluate the effectiveness of the suggested and current models are succinctly summarized below.

3.1 Accuracy

It is the number of appropriately categorized signs to the entire number of signs employed for the SL categorization. It clearly indicates the prediction classifier being trained correctly to perform generally.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

Here, True Positive (TP) - If the model properly identified the sign as the exact class; True Negative (TN) - When a model accurately predicts a sign that does not belong to a given class, False positive (FP) - When the actual class of sign is false but the projected class is true. False negative (FN) - When the true class of a sign is expected to be false.

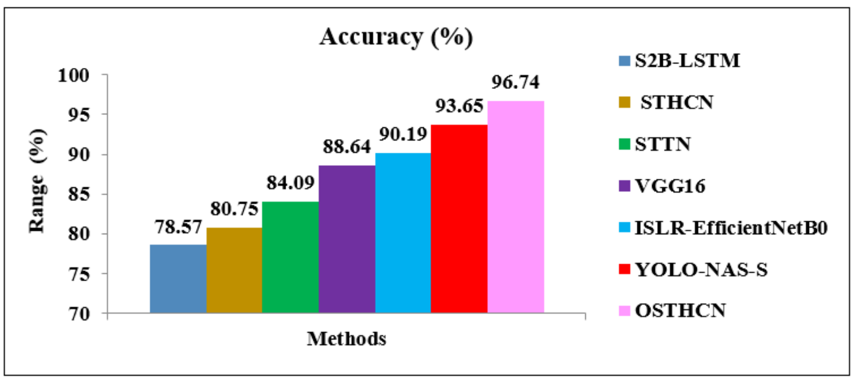

Fig 6. Accuracy Comparison for proposed and existing models

In Figure 6, the impact of the existing and proposed model on ISL-CSLRT data is drawn in terms of prediction accuracy. It shows that the proposed OSTHCN algorithm achieves accuracy compared to the other algorithms due to the use of various features and annotations, which are learned automatically and without difficulty. On average, the OSTHCN algorithm increases the accuracy by 23.13%, 19.80%, 15.04%, 9.14%, 7.26% and 3.29% in contrast with the S2B-LSTM, STHCN, STTN, VGG16, ISLR-EfficientNetB0 and YOLO-NAS-S respectively.

3.2 Precision

The precision used in SLR task to assess model performance primarily provides information about the number of FP in the dataset. Lower amount of false positive values and higher accuracy score.

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

Figure 7 portrays the precision values obtained by the different algorithms for predicting ISL performance in terms of precision. It is shown that the OSTHCN algorithm can enhance the precision values by 20.13%, 16.87%, 13.89%, 10.29%, 6.51%, and 2.76% respectively, compared to the S2B-LSTM, STHCN, STTN, VGG16, ISLR-EfficientNetB0 and YOLO-NAS-S algorithms, accordingly. Thus, it is concluded that the OSTHCN algorithm enhances the prediction performance by learning all the factors about ISL information in contrast with the existing algorithms.

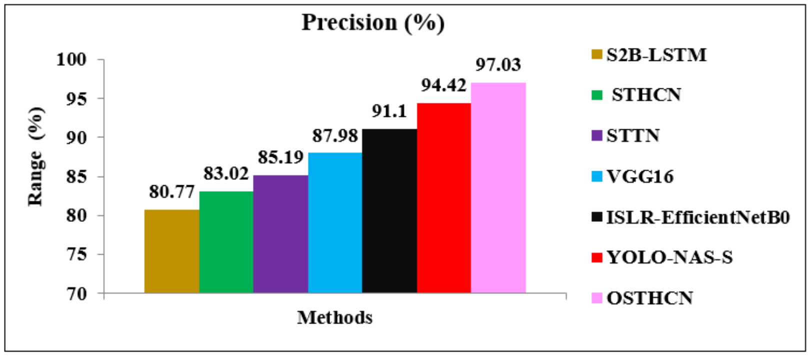

Fig 7. Precision Comparison for proposed and existing models

3.3 Recall

The recall score is used in SLR to evaluate model performance; it primarily informs about FN available in the dataset. A higher recall score means fewer FN values.

$$Recall = \frac{TP}{TP+FN} \tag{7}$$

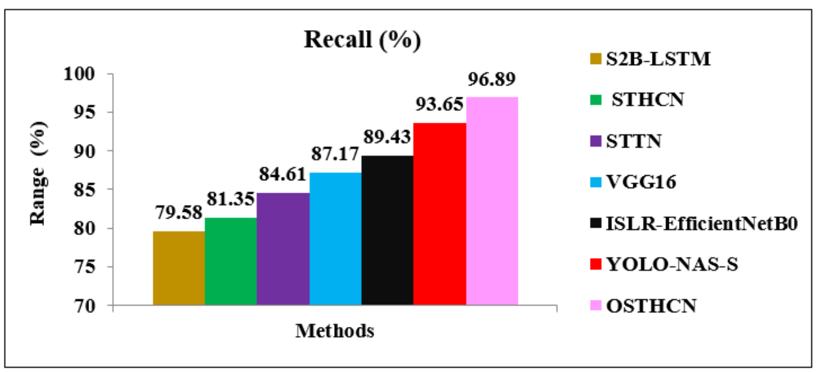

Fig 8. Recall Comparison for proposed and existing models

Figure 8 illustrates the effect of the current model and the suggested model on the ISL-CSLRT data in terms of recall values. It should be highlighted that, in comparison to other algorithms already in use, the suggested OSTHCN algorithm may effectively maximize the recall of forecasting ISL performance by taking all factors into account. In comparison to S2B-LSTM, STHCN, STTN, VGG16, ISLR-EfficientNetB0 and YOLO-NAS-S algorithms improves accuracy on average by 21.75%, 19.10%, 14.51%, 11.15%, 8.34% and 3.46%, correspondingly.

3.4 F1-Score

It is the average of the two measures of precision and recall

$$F1-score = 2\times\frac{Precision \bullet Recall}{Precision+Recall} \tag{8}$$

Figure 9 plots the F1-Score of different ISL prediction algorithms in terms of F1-Score on ISL-CSLRT dataset. It is noted that the proposed OSTHCN algorithm can efficiently maximize the F1-Score of predicting ISL performance by considering all criteria, compared to other existing algorithms. The OSTHCN algorithm increases F1-Score by 22.47%, 20.65%, 15.34%, 11.95%, 8.27% and 29.2% in contrast with the S2B-LSTM, STHCN, STTN, VGG16, ISLR-EfficientNetB0 and YOLO-NAS-S respectively.

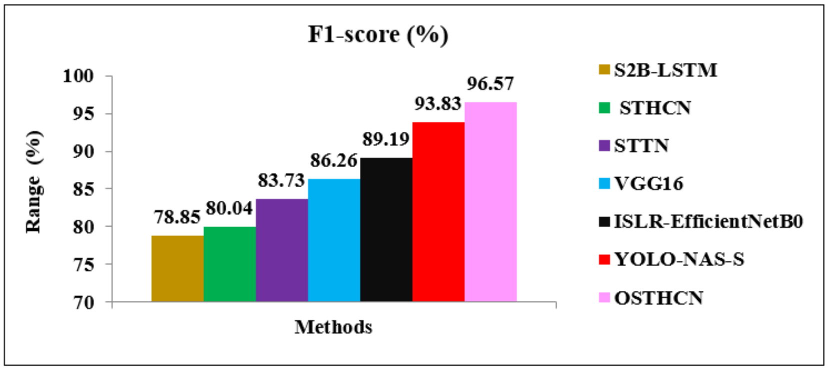

Fig 9. F1-score Comparison for proposed and existing models

From the above analysis, it is determined that the proposed OSTHCN model achieves greater performance than other existing SLR models. This is because the proposed model effectively tackles occlusion challenges in skeleton extraction for ISL recognition and translation. It addresses the disconnection of subgraphs caused by occlusions and identifies these clusters using the shortest cost paths. This allows for accurate prediction of the complete skeleton, leading to significant improvements in ISL recognition and translation performance.

4 Conclusion

In this study, OSTHCN model is proposed for ISL recognition and translation to effectively address the occlusions in human skeleton extraction using a skeleton occupancy likelihood map estimated through B-Spline curves. This method creates a precise 3D skeleton model by predicting occluded structures and updating the probability distribution based on branch clusters and minimum-cost path searches. The developed OSTHCN model achieves 96.74% of accuracy on ISL-CSLTR dataset compared to existing models. This model features a distinctive occlusion-handling strategy for skeleton extraction, estimating occluded parts and using minimal cost route searches to integrate fragmented subgraphs, resulting in a more precise skeleton model. Its dynamic subgraph integration is ideal for live translation systems and interactive sign language interfaces, enhancing assistive technologies and communication tools. However, manually assigned hyper-parameters in sequence learning led to high computational complexity and reduced accuracy. Future improvements will involve advanced optimization models for more efficient ISL detection.

References

1) Joksimoski B, Zdravevski E, Lameski P, Pires IM, Melero FJ, Martinez TP, et al. Technological solutions for sign language recognition: a scoping review of research trends, challenges and opportunities. *IEEE Access*. 2022;10:40979–40998. Available from: https://ieeexplore.ieee.org/document/9739689/authors#authors.
2) Ibrahim NB, Zayed HH, Selim MM. Advances, challenges and opportunities in continuous sign language recognition. *Journal of Engineering and Applied Sciences*. 2020;15(5):1205–1227. Available from: https://www.researchgate.net/publication/337657675_Advances_Challenges_and_Opportunities_in_Continuous_Sign_Language_Recognition.
3) Papastratis I, Dimitropoulos K, Daras P. Continuous sign language recognition through a context-aware generative adversarial network. *Sensors*. 2021;21(7):1–20. Available from: https://doi.org/10.3390/s21072437.
4) Kamble N, More N, Wargantiwar O, More S. Deep Learning-Based Sign Language Recognition and Translation. In: G R, Y EA, S P, editors. Soft Computing for Security Applications. ICSCS 2023. Advances in Intelligent Systems and Computing;vol. 1449. Springer. 2023;p. 49–63. Available from: https://doi.org/10.1007/978-981-99-3608-3_4.
5) Kaur H, Bedi SK, Rao KM, Lekhana MA. Deep Neural Networks for Image-Based Indian Sign Language Recognition: A Comprehensive Review and Practical Analysis. In: S D, S S, CAC C, H R, JC B, editors. Advances in Data-Driven Computing and Intelligent Systems. ADCIS 2023. Lecture Notes in Networks and Systems;vol. 890. Springer. 2024;p. 49–64. Available from: https://link.springer.com/chapter/10.1007/978-981-99-9531-8_5.
6) Sharma S, Singh SA. Spatio-Temporal Framework for Dynamic Indian Sign Language Recognition. *Wireless Personal Communications*. 2023;132:2527–2541. Available from: https://doi.org/10.1007/s11277-023-10730-8.
7) Muthusamy P, Murugan GP. Recognition of Indian Continuous Sign Language Using Spatio-Temporal Hybrid Cue Network. *International Journal of Intelligent Engineering & Systems*. 2023;16(6):874–885. Available from: https://oaji.net/articles/2023/3603-1698222905.pdf.
8) Cui Z, Zhang W, Li Z, Wang Z. Spatial-temporal transformer for end-to-end sign language recognition. *Complex & Intelligent Systems*. 2023;9(4):4645–4656. Available from: https://link.springer.com/article/10.1007/s40747-023-00977-w.
9) Sharma P, Balpande S, Sharma N, Gourshete GD, Bendale D, Kalme GR. Indian sign language recognition and translation: Text to sign language using deep learning technique. *AIP Conference Proceedings*. 2024;3112. Available from: https://pubs.aip.org/aip/acp/article-abstract/3112/1/020027/3295724/

Indian-sign-language-recognition-and-translation?redirectedFrom=fulltext.
10) Vidhyalakshmi MK, Jagadeesh K, Dharmaraj BS, Sankar NP, Kumar RG. Indian Sign Language Recognition using transfer learning with Efficient Net. *Smart Science*. 2024;12(2):269–280. Available from: https://doi.org/10.1080/23080477.2024.2320956.
11) Priya K, Sandesh BJ. Developing an offline and real-time Indian sign language recognition system with machine learning and deep learning. *SN Computer Science*. 2024;5(3):1–12. Available from: https://link.springer.com/article/10.1007/s42979-023-02482-w.
12) Elakkiya R, Natarajan B. ISL-CSLTR: Indian Sign Language Dataset for Continuous Sign Language Translation and Recognition. 2021. Available from: https://doi.org/10.17632/kcmpdxky7p.1.