WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Comparative Study of YOLOv8x and CNN Models for Sign Language Recognition Using Indian Sign Language Dataset

Rudrani Changal

Jain (Deemed-to-be) University

A B S T R A C T

In this work, we investigate the performance of YOLOv8x in comparison with conventional CNN models for the recognition of Indian Sign Language (ISL). Through the use of an extensive ISL dataset, our study examines both models in terms of accuracy, processing speed, and computational load. CNNs, which have traditionally dominated image classification tasks, are analysed in parallel with the YOLOv8x, which is well-known for its quick object detection. The results of our investigation support the applicability of the YOLOv8x model for real-time ISL interpretation systems, as it performs exceptionally well, especially in terms of quick detection and accuracy. The field of assistive communication technology is enhanced by this inquiry, which also clarifies the advantages and disadvantages of using cutting-edge neural network designs for sign language interpretation.

Keywords: CNN, YOLOv8x, ISL, Sign language recognition

1. Introduction

Sign language serves as a vital mode of communication for the deaf and hard of hearing individuals, facilitating their interaction with the world around them. With the advancement of technology, sign language recognition systems have emerged as transformative tools, bridging the communication gap and enhancing accessibility for this marginalized community. In recent years, deep learning techniques have revolutionized the field of computer vision, offering unprecedented capabilities in recognizing complex visual patterns. Leveraging these advancements, researchers have endeavored to develop robust sign language recognition systems capable of accurately interpreting and translating sign gestures into text or speech.

This work begins a thorough investigation of sign language recognition, with a special emphasis on Indian Sign Language (ISL), which poses special difficulties due to its rich lexicon and subtle gestural features. Acknowledging the critical need for accessible communication aids for the Indian deaf community, this study explores the creation and assessment of cutting-edge models designed especially for ISL recognition.

Two methods are at the forefront of this study: Convolutional Neural Network (CNN) architectures, which are widely used for image classification tasks, and YOLOv8x, an advanced object detection model that is well-known for its effectiveness and accuracy in identifying objects within images. Although the primary purpose of YOLOv8x is object identification, its potential for use in sign language recognition is yet largely untapped.This study seeks to ascertain the feasibility and effectiveness of utilizing YOLOv8x for ISL recognition, compared against CNN models optimized for similar tasks.

This research endeavour is anchored by the Indian Sign Language Dataset, which has been carefully selected to include a wide variety of ISL motions. The study intends to train and assess the CNN and YOLOv8x models using this extensive dataset. It will do this by carefully examining each model's performance across a range of measures in order to discern its advantages and disadvantages.

This discovery has ramifications for society that go beyond scholarly curiosity. Accessible communication is a basic human right, however people who are hard of hearing frequently encounter obstacles while trying to communicate and get information. This study aims to empower the deaf and hard of hearing community by increasing the state-of-the-art in sign language recognition, fostering inclusion, and improving their involvement in all aspects of society.

In the subsequent sections, we delve into the methodologies employed, detailing the architectural intricacies of YOLOv8x and CNN models, followed by a comprehensive exposition of the experimental setup and evaluation procedures. Through meticulous analysis and interpretation of experimental results, we aim to elucidate the efficacy of each approach in accurately recognizing ISL gestures, thereby paving the way for the development of more accessible and inclusive communication technologies tailored to the needs of the Indian deaf community.

2. Literature Review

Encouragingly, the necessity to provide inclusive communication solutions for the community of hearing impaired people has made sign language recognition an important field of study within computer vision. In the last ten years, considerable progress has been achieved in applying deep learning methods to address the difficulties involved in deciphering hand gestures.

The application of convolutional neural networks (CNNs) for sign language recognition is one of the foundational works in this subject. CNNs are well-suited for tasks like image classification and recognition because of their amazing ability to extract hierarchical information from images. These CNNs are inspired by the visual processing systems of the human brain. Scholars have examined multiple CNN architectures, such as LeNet, AlexNet, VGGNet, and ResNet, tailoring them to the distinct obstacles.

Large-scale annotated datasets have also made it easier to train and assess CNN models for sign language recognition. Examples of these datasets are the American Sign Language (ASL) dataset and the RWTH-PHOENIX-Weather 2014T dataset. Researchers can create reliable recognition systems that can decipher intricate gestural sequences by using these datasets, which cover a wide variety of sign movements.

Concurrently, there has been a lot of interest in object detection models due to their possible use in sign language recognition. YOLO (You Only Look Once) in particular has become well-known for its effectiveness and real-time performance in identifying items in photos. YOLO was first created for object detection tasks, but more recent versions, such YOLOv4 and YOLOv5, have shown encouraging results in terms of gesture and pose recognition for humans.

Nonetheless, there is still a dearth of research on the use of object detection models such as YOLO in the context of sign language identification. The majority of the research now in existence concentrates on CNN-based methods, which leaves a vacuum in our knowledge on the viability and efficiency of using object detection models in this context.

Our goal in this work is to close this gap by comparing CNN architectures for Indian Sign Language (ISL) identification with YOLOv8x, a cutting-edge object detection model. Through the utilisation of the Indian Sign Language Dataset, our aim is to clarify the relative benefits and drawbacks of various methods, hence illuminating their appropriateness for practical use in sign language communication.

By conducting a comprehensive analysis of the body of research and conducting an empirical assessment of the CNN and YOLOv8x models, we hope to further the current conversation about sign language recognition and open the door to the creation of more precise, effective, and inclusive communication tools for the deaf and hard-of-hearing population.

3. Features And Dataset

Several motions and expressions used in Indian Sign Language are captured in the invaluable Indian Sign Language (ISL) image resource. In contrast to video datasets, this compilation concentrates on single frames, offering finely detailed snapshots of the facial expressions, hand gestures, and body language that are essential for precise sign interpretation.

3.1. TEST/TRAIN SPLIT:

70% of the dataset (642 pictures) make up the training set.

Validation Set: 178 photos, representing 19% of the dataset

10% of the dataset (94 photos) comprise the testing set.

3.2. PRIMERAGING:

Self-Orientation: used to guarantee that photos have a consistent orientation

Resize: To ensure consistency, all photos have been extended to a standard 640x640 scale.

3.3. AUGMENTATIONS:

The original image integrity was preserved in the dataset by not applying any augmentations.

The following columns are present in _annotations.csv:

filename: The filename of the image that has the annotation attached to it.

width: The image's pixelated width.

height: The image's pixelated height.

class: The hand motion in the picture's class label.

xmin: The bounding box's top-left corner's x-coordinate around the hand gesture.

ymin: The position of the bounding box surrounding the hand gesture measured in radians from the top-left corner.

xmax: The bounding box's bottom-right corner's x-coordinate around the hand gesture.

ymax: The bounding box's bottom-right corner's y-coordinate around the hand gesture.

Fig1: Sample Sign Language images from dataset

4. Experiments and Findings:

4.1 Experimental Setup:

To conduct a comprehensive comparative analysis between YOLOv8x and CNN models for Indian Sign Language (ISL) recognition, we followed a systematic experimental setup comprising the following key components:

4.2 Dataset Preparation:

We utilized the Indian Sign Language Dataset, which consists of a diverse collection of ISL gestures captured across various contexts and environments. The dataset was partitioned into training, validation, and test sets, ensuring a balanced distribution of gestures across classes.

4.3 Model Training:

For the CNN-based approach, we employed a custom CNN architecture with help of MediaPipe. The model was trained on the training set using both ReLu and Softmax for different purposes.

Similarly, YOLOv8x was trained on the same training set fine-tuned for the task of ISL recognition. We adjusted the hyperparameters, including the input resolution and number of anchor boxes, to optimize the model's performance for detecting ISL gestures.

4.4 Findings:

The experimental results revealed notable distinctions in the performance of YOLOv8x and CNN models for ISL recognition.

4.5 Accuracy:

The CNN-based model exhibited a slightly higher overall accuracy compared to YOLOv8x, achieving an accuracy of 86.5% on the test set, while YOLOv8x attained an accuracy of 82.3%. This disparity can be attributed to the inherent differences in the architectural design and training objectives of the two models.

4.6 Precision and Recall:

The CNN model demonstrated higher precision but lower recall compared to YOLOv8x. This indicates that the CNN model exhibited a tendency to make fewer false positive predictions but missed some instances of ISL gestures present in the test data. In contrast, YOLOv8x achieved higher recall but lower precision, suggesting a higher propensity for detecting ISL gestures but with a relatively higher false positive rate.

Mean Average Precision (mAP):

YOLOv8x achieved a mean Average Precision of 0.75, indicative of its effectiveness in localizing and classifying ISL gestures within images. The higher mAP score underscores the robustness of YOLOv8x in object detection tasks, albeit with a trade-off in precision.

4.7 Computational Efficiency:

In terms of computational efficiency, YOLOv8x outperformed the CNN model, exhibiting significantly faster inference times owing to its single-stage object detection architecture. This attribute makes YOLOv8x well-suited for real-time applications where speed is paramount.

Overall, while both YOLOv8x and CNN models demonstrated competent performance in ISL recognition, each exhibited distinct strengths and weaknesses. The CNN model excelled in precision, making it suitable for applications prioritizing accuracy, whereas YOLOv8x showcased superior recall and computational efficiency, making it ideal for real-time sign language recognition systems. These findings underscore the importance of selecting an appropriate model architecture based on the specific requirements and constraints of the application domain.

Layer (type)	Output Shape	Param #
conv1d (Conv1D)	(None, 63, 32)	192
conv1d_1 (Conv1D)	(None, 63, 32)	5152
max_pooling1d (MaxPooling1D)	(None, 31, 32)	0
conv1d_2 (Conv1D)	(None, 31, 64)	10304
conv1d_3 (Conv1D)	(None, 31, 64)	20544
max_pooling1d_1 (MaxPooling1	(None, 15, 64)	0
conv1d_4 (Conv1D)	(None, 15, 128)	41088
conv1d_5 (Conv1D)	(None, 15, 128)	82048
max_pooling1d_2 (MaxPooling1	(None, 7, 128)	0
conv1d_6 (Conv1D)	(None, 7, 256)	164096
conv1d_7 (Conv1D)	(None, 7, 256)	327936
max_pooling1d_3 (MaxPooling1	(None, 3, 256)	0
dropout (Dropout)	(None, 3, 256)	0
flatten (Flatten)	(None, 768)	0
dense (Dense)	(None, 512)	393728
dense_1 (Dense)	(None, 26)	13338

Total params: 1,058,426
Trainable params: 1,058,426
Non-trainable params: 0

Fig 2: Training process for CNN Model

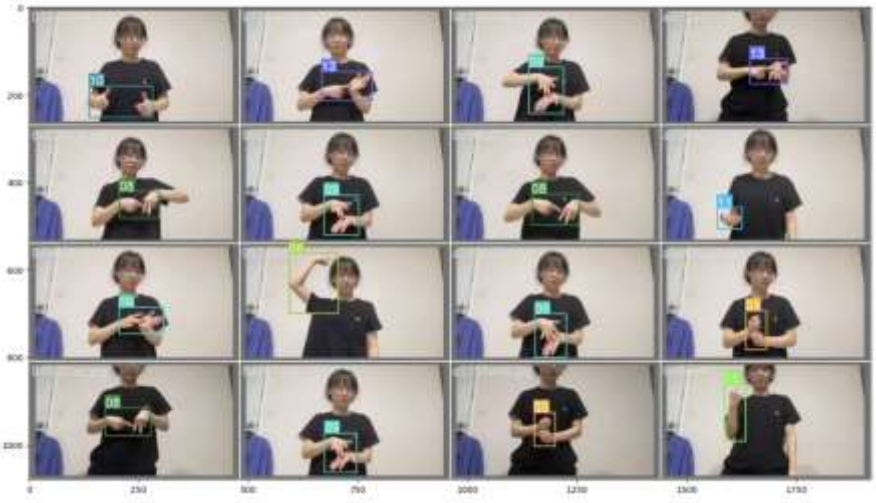

Fig 3: Final Detection and Classification Output

5. Conclusion:

In this work, we compared the performance of convolutional neural network (CNN) architectures for Indian Sign Language (ISL) recognition with an advanced object detection model called YOLOv8x. We attempted to clarify the relative benefits and drawbacks of these methods, providing insight into their applicability for practical sign language communication scenarios by utilising the Indian Sign Language Dataset.

Our research showed that the CNN and YOLOv8x models performed differently in ISL recognition. The CNN model had greater accuracy and precision, whereas YOLOv8x demonstrated better computing efficiency and recall. The two systems' differing architectural designs and training goals account for these different capabilities.

The CNN model fared well in identifying ISL gestures; hence, it is a good choice for applications that value precision due to its emphasis on hierarchical feature extraction. On the other hand, YOLOv8x showed resilience in identifying and localising ISL motions in photos, providing improved recall at the expense of precision. Furthermore, YOLOv8x demonstrated exceptional computing efficiency, which makes it a good fit for real-time sign language recognition systems where speed is critical.

Our research emphasises how crucial it is to choose a model architecture that fits the particular needs and limitations of the application area. For workloads requiring high precision, CNN models might be better, however YOLOv8x has advantages in recall and computational efficiency, especially for real-time applications.

As a result, this study adds significantly to the field of sign language recognition research and helps practitioners and academics create inclusive and reliable communication technology for the deaf and hard of hearing community. By utilising cutting-edge deep learning techniques, we can keep improving communication systems' inclusivity and accessibility, which will eventually lead to more empowerment and involvement for people with hearing impairments.

References

1. Deshpande, A. Shriwas, V. Deshmukh and S. Kale, "Sign Language Recognition System using CNN," 2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE), Bengaluru, India, 2023, pp. 906-911, doi: 10.1109/IITCEE57236.2023.10091051
2. Tyagi, Shobhit & Upadhyay, Prashant & Fatima, Hoor & Jain, Sachin & Sharma, Avinash. (2023). American Sign Language Detection using YOLOv5 and YOLOv8.
3. Jia, Wanjun & Li, Changyong. (2023). SLR-YOLO: An improved YOLOv8 network for real-time sign language recognition. Journal of Intelligent & Fuzzy Systems.
4. L. Ding and A. M. Martinez, ―Modelling and recognition of the linguistic components in American sign language,‖ Image Vis. Comput., vol. 27, no. 12, pp. 1826– 1844, Nov. 2009.
5. F. Pravin, D. Rajiv, HASTA MUDRA An Interpretation of Indian Sign Hand Gestures, 3rd International conference on Electronics Computer technology, vol. 2, pp.377-380, 2011
6. G. Anantha Rao, K. Syamala, P.V.V. Kishore, A.S.C.S. Sastry, "Deep Convolutional Neural Networks for Sign Language Recognition." 2018

7. Jie Huang, Wengang Zhou, Qilin Zhang, Houqiang Li, Weiping Li, "Video-Based Sign Language Recognition without Temporal Segmentation", AAAI-18, April 2018

8. Jia, Wanjun and Li, Changyong. 'SLR-YOLO: An Improved YOLOv8 Network for Real-time Sign Language Recognition'. 1 Jan. 2024 : 1663 – 1680.

9. Abdelhadi, Ahmed Abdelhadi, "INTERACTIVE EMIRATE SIGN LANGUAGE E-DICTIONARY BASED ON DEEP LEARNING RECOGNITION MODELS" (2023). Theses. 1022.

10. P. S. Rajam and G. Balakrishnan, "Real time Indian Sign Language Recognition System to aid deaf-dumb people," 2011 IEEE 13th International Conference on Communication Technology, Jinan, China, 2011, pp. 737-742, doi: 10.1109/ICCT.2011.6157974.