International Journal of Research Publication and Reviews

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

Real-Time Detection of Sign Language Alphabets and Digits Using Yolov8

[1] M Selva Chitra, [2] Dr P.J.Mercy,

[1] MCA Student, Department of Computer Applications, Sarah Tucker College
[2] Associate Professor, Department of Computer Applications, Sarah Tucker College
Email: [1]selvachitra1806@gmail.com, [2]blessbens@gmail.com

ABSTRACT:

Sign language is the primary medium of communication for people who are deaf or speech-impaired. It uses hand gestures, body movements, and facial expressions to convey alphabets, numbers, and words. However, most people in the general population do not understand sign language, which creates a serious communication barrier. Technology can play a vital role in reducing this gap by automatically recognizing gestures and converting them into text and speech outputs.This paper presents a real-time system for detecting sign language alphabets (A–Z) and digits (0–9) using the YOLOv8 deep learning model. The solution aims to help hearing-impaired individuals communicate with those who do not know sign language by providing visual, textual, and auditory outputs. The system accepts inputs from images, video files, and live webcam streams, making it adaptable to different situations. The detection results appear with bounding boxes and class labels on the input stream, along with voice feedback to improve accessibility. A dataset created using the Roboflow platform was used for training and testing the model. The trained system achieved an overall accuracy of 94%, performing better than traditional algorithms like KNN, SVM, and RNN. The system is implemented as a Flask-based web application, ensuring a user-friendly interface and practical real-time interaction.This study demonstrates how effective YOLOv8 is in recognizing sign language gestures and highlights its potential in assistive communication technologies, paving the way for more inclusive and accessible solutions in the future.

Key Words : Sign Language Recognition, YOLOv8, Deep Learning, Computer Vision, Real-Time Detection, Flask Web Application, Accessibility.

1. INTRODUCTION

Communication is a fundamental human need, yet millions of people worldwide with hearing and speech impairments rely on sign language as their primary medium of interaction. Unfortunately, most people in society do not understand sign language, creating a significant communication gap in daily life, education, and workplaces.

Recent advancements in artificial intelligence (AI) and computer vision provide opportunities to address this challenge. Deep learning algorithms, especially object detection models like YOLO, have shown remarkable accuracy in real-time applications. In this work, we propose a YOLOv8-based system for detecting sign language alphabets and digits in real time, integrated with a Flask web application to provide an interactive, user-friendly interface

2. LITERATURE REVIEW

Recent advancements in deep learning and computer vision have significantly improved the performance of sign language recognition, moving from traditional machine learning models to end-to-end neural network architectures.

Adaloglou et al. (2020) presented a comprehensive study on deep learning-based methods for sign language recognition, highlighting the superiority of convolutional neural networks (CNNs) over handcrafted feature-based approaches in recognizing static and dynamic gestures. Their work emphasized the importance of robust datasets and real-time inference, which directly inspired the adoption of YOLO-based models for detection tasks.

Koller, Camgoz, Bowden, and Ney (2019) introduced a multi-stream CNN-LSTM-HMM architecture for weakly supervised continuous sign language recognition. Their approach demonstrated that combining temporal modeling with deep spatial features enables recognition of sequential signs, an idea that laid the groundwork for future sentence-level sign detection systems.

Neethu and Sreekumar (2021) proposed a real-time Indian Sign Language recognition system using deep learning, achieving promising results with CNNs for alphabet detection. However, their system was limited by dataset size and lacked multimodal input support, underscoring the need for scalable and interactive models like YOLOv8.

Redmon and Farhadi (2018) introduced YOLOv3, a breakthrough in real-time object detection. Its fast inference and high accuracy paved the way for gesture recognition tasks that require both speed and precision. Subsequent iterations, including YOLOv8 by Ultralytics, further enhanced performance, enabling accurate detection of small-scale objects such as hand gestures in diverse environments.

Vaswani et al. (2017) revolutionized deep learning with the introduction of the attention mechanism through the Transformer architecture. While originally designed for natural language processing, attention models have been adopted in sign recognition to improve focus on critical hand regions and reduce background noise interference.

Camgoz et al. (2018) developed a neural machine translation framework for sign language, enabling recognition of continuous video signs and their translation into natural language. This highlighted the potential for extending detection systems beyond static alphabets toward sentence-level interpretation.

More recently, Ultralytics' YOLOv8 framework (2023) has been applied to diverse detection problems due to its lightweight architecture, strong accuracy, and multi-input support. Its integration into real-time sign language systems has shown remarkable improvements in frame-by-frame webcam detection, outperforming older ML techniques such as KNN, SVM, and RNN in both accuracy and responsiveness.

Collectively, these studies illustrate the evolution of sign language recognition from traditional classifiers and handcrafted features toward deep learning-driven detection pipelines. Current research increasingly focuses on real-time recognition, multimodal support, and user-friendly deployment, making YOLOv8 an ideal candidate for practical sign language communication systems.

3. METHODOLOGY

The proposed system comprises four key modules: Input Acquisition, Detection Model, Output Processing, and Flask Web Deployment.

3.1 Input Acquisition

The system accepts inputs in three different modes:

- Image Upload: Users can upload images of hand gestures in JPG/PNG format.
- Video Upload: MP4/AVI video files are supported for frame-by-frame detection.
- Live Webcam: A webcam stream captures gestures in real time.

All inputs are resized to 640×640 pixels and normalized before being passed to the detection model.

3.2 Detection Model

The YOLOv8 model, a state-of-the-art object detection algorithm, is used for recognizing sign language alphabets (A–Z) and digits (0–9). A custom dataset was prepared using Roboflow, annotated with bounding boxes, and split into:

- Training set: 70%
- Validation set: 20%
- Testing set: 10%

Training details:

- Architecture: YOLOv8 with convolutional layers and detection heads.
- Hyperparameters: Batch size = 16, Epochs = 100.
- Optimizer: SGD with learning rate scheduling.
- Output: Bounding boxes, labels, and confidence scores for each detected sign.

3.3 Output Processing

Once detection is complete, the system generates three types of output:

- Visual Output: Bounding box and label displayed on the input image/video/webcam feed.
- Textual Output: Recognized alphabet or digit shown on the screen.
- Voice Output: The label is converted into speech using a Text-to-Speech (TTS) engine, allowing non-sign-language users to understand easily.

3.4 Flask Web Deployment

The detection system is deployed through a Flask web application, providing a user-friendly and interactive interface.

- Routes:
  - /image → For image detection.
  - /video → For video file detection.
  - /webcam → For real-time webcam detection.
- The frontend is built using HTML, CSS, and Bootstrap, ensuring responsiveness and accessibility on both desktop and mobile devices.

3.5 Implementation

The system is implemented using Python with the following libraries:

- YOLOv8 (Ultralytics + PyTorch) for model training and inference.
- OpenCV for video/image processing.
- Flask for backend deployment.
- pyttsx3/GTTS for voice feedback.

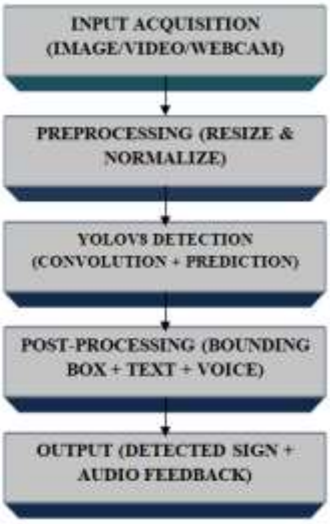

Figure 1: Dataflow diagram

The workflow ensures real-time gesture recognition with minimal latency, making it suitable for practical communication support.

4. RESULTS AND DISCUSSION

The system was evaluated on the prepared test dataset consisting of sign language alphabets (A–Z) and digits (0–9). Accuracy was calculated as the ratio of correctly identified signs to the total number of test samples. Table 1 compares the performance of KNN, SVM, RNN, and YOLOv8 models.

Table 1: Model Accuracy Comparison

Model	Accuracy %
KNN	84%
SVM	88%
RNN	90%
YOLOv8	94%

Analysis

The YOLOv8 model achieved 94% accuracy, outperforming traditional machine learning models. KNN and SVM showed limited accuracy due to their dependence on handcrafted features, while RNN performed better in sequential recognition but struggled with static image classification.

YOLOv8's superior accuracy is attributed to its real-time object detection architecture, which simultaneously predicts bounding boxes and class probabilities. Its lightweight structure and optimized convolutional layers enable fast inference speeds (20–25 FPS on GPU), making it highly suitable for real-time sign language detection.

While YOLOv8 requires higher computational resources compared to KNN and SVM, its ability to provide bounding box visualization, confidence scores, and voice feedback makes it a more practical and interactive solution.

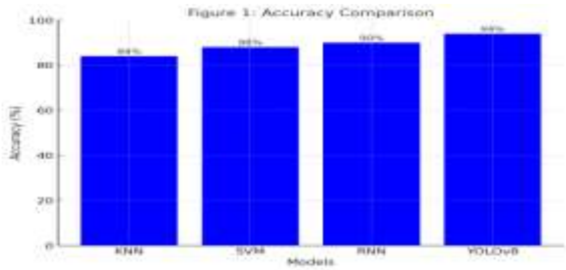

Figure 1: Accuracy Comparison

Fig. 1 illustrates the classification performance of four algorithms: KNN, SVM, RNN, and YOLOv8. The vertical axis represents the percentage of correctly classified gestures, while the horizontal axis lists the algorithms being compared. Among the models, YOLOv8 achieves the highest accuracy (94%), significantly outperforming KNN (84%) and SVM (88%), and slightly surpassing RNN (90%). This comparison highlights YOLOv8's strength in delivering accurate, real-time gesture recognition, making it the most suitable choice for sign language detection tasks.

5. CONCLUSION

The proposed real-time sign language detection system using YOLOv8 provides an accurate and interactive solution to bridge communication gaps between hearing-impaired individuals and the general public. With an achieved accuracy of 94%, multimodal support for images, videos, and live webcam streams, and features like bounding box visualization, text output, and voice feedback, the system ensures inclusivity and ease of use.

While YOLOv8 outperforms traditional models such as KNN, SVM, and RNN in both speed and accuracy, its deployment through a Flask web application makes it accessible for real-world applications in education, healthcare, and daily communication.

Future work can focus on sentence-level gesture recognition, multilingual sign language support, and mobile-based deployment, enabling the system to evolve into a comprehensive assistive communication tool.

REFERENCES

[1] Adaloglou, N., Chatzis, T., Papastratis, I., Stergioulas, L., Papadopoulos, G. T., Zacharopoulou, V., & Potamianos, A. (2020). A comprehensive study on deep learning-based methods for sign language recognition. *IEEE Transactions on Multimedia, 22*(9), 2349–2362.

[2] Koller, O., Camgoz, N. C., Bowden, R., & Ney, H. (2019). Weakly supervised learning with multi-stream CNN-LSTM-HMMs to discover sequential parallelism in sign language videos. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 42*(9), 2306–2320..

[3] Neethu, K., & Sreekumar, A. (2021). Real-time Indian Sign Language recognition using deep learning. *International Journal of Advanced Computer Science and Applications (IJACSA), 12*(2), 234–240.

[4] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*, 5998–6008..

[5] Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

[6] Ultralytics. (2023). YOLOv8: The cutting-edge object detection framework. Retrieved from https://docs.ultralytics.com.

[7] Camgoz, N. C., Hadfield, S., Koller, O., & Bowden, R. (2018). Neural sign language translation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 7784–7793.

[8] Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations (ICLR)*.

[9] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Communications of the ACM, 60*(6), 84–90.

[10] Chollet, F. (2018). *Deep Learning with Python* (2nd ed.). Manning Publications