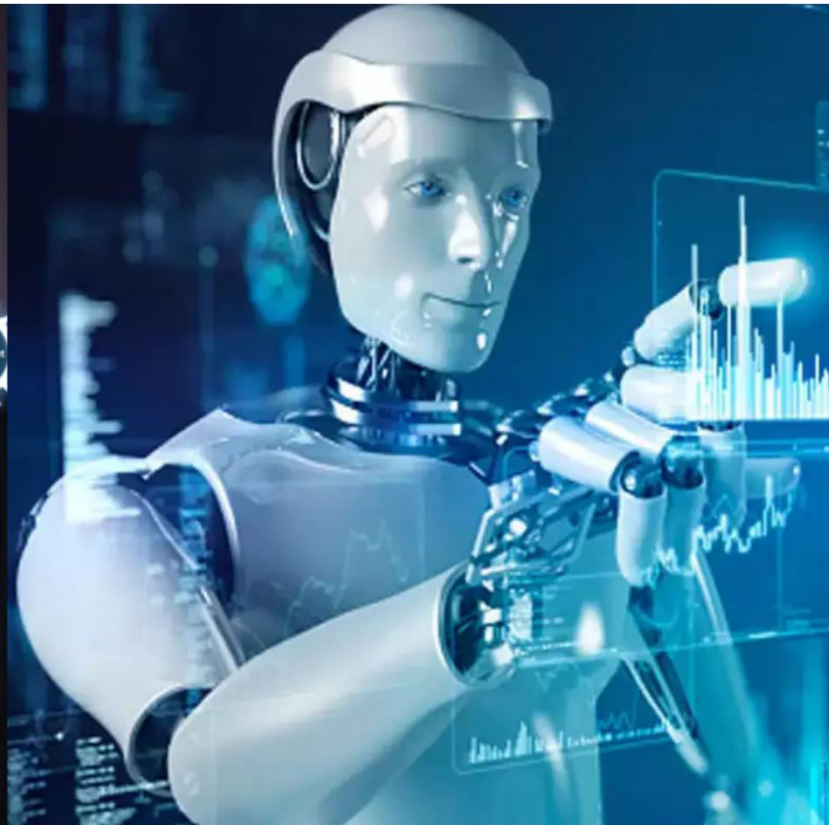
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 15, Issue 3, March 2026

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503096|

Silent Sentinels: A Multimodal AI Framework for Gaze-Based Communication and Assistance in Locked-In Syndrome

R. Subashini[1], Dhivyashree S[2], Lakchaya V[3], Mythili S[4]

Assistant Professor, Department of Artificial Intelligence and Data Science, Kongunadu College of Engineering and Technology, Trichy, India[1]

Department of Artificial Intelligence and Data Science, Kongunadu College of Engineering and Technology, Trichy, India[2,3,4]

ABSTRACT: Locked-In Syndrome (LIS) is a severe neurological condition in which individuals experience near-total loss of voluntary motor control while retaining mental awareness, making communication extremely difficult. In order to overcome this issue, a gaze-based interaction system enables hands-free communication by interpreting eye movements captured through a webcam. The proposed system employs MediaPipe and OpenCV to track eye landmarks and iris movements. Gaze-based text selection is combined with eye-blink based selection to support reliable interaction. Integration of word prediction and personalized phrase suggestions reduces the effort required for communication. Selected text is converted into speech using text-to-speech (TTS), which enables local voice alerts for immediate caregiver awareness. Additionally, email alerts using SMTP protocols are sent to facilitate remote caregiver communication when assistance is needed. Fatigue detection with pause mode and error correction with undo selection minimize unintended inputs and improve user interaction. This solution can be integrated in assistive care units and rehabilitation centers to support individuals with Locked-In Syndrome. Through a comprehensive review of state-of-the-art research from 2021 to 2026, this paper demonstrates that deep learning-based blink detection achieves superior accuracy compared to traditional methods , while multi-stage hierarchical gaze selection significantly improves text entry speed and usability . The integration of large language models for word prediction has been shown to achieve keystroke savings of up to 70% . This paper presents a unified framework synthesizing these advances into a practical, deployable system for LIS communication.

KEYWORDS: Locked-In Syndrome, Gaze-Based Interaction System, Eye-Blink Detection, OpenCV, MediaPipe, Text-to-Speech, Word Prediction, Assistive Technology, Human-Computer Interaction, Deep Learning.

I. INTRODUCTION

Locked-In Syndrome (LIS) is one of the most severe physical disabilities. LIS was first identified by Plum and Posner in 1966. It is defined by quadriplegia and anarthria (loss of speech) with retained consciousness and intellect. Sufferers of LIS are conscious, alert, and mentally sound but paralyzed and speechless due to damage to the ventral pons in the brain. They can only communicate through vertical eye movements and blinking, which are the only voluntary motor functions that remain intact. The sense of isolation that these patients experience has been aptly captured in Jean-Dominique Bauby's biography "The Diving Bell and the Butterfly," which was written by eye blinking [1].

The number of cases of diseases that can cause LIS or other forms of severe motor impairment is considerable. Amyotrophic lateral sclerosis, brainstem strokes, severe muscular dystrophy, and cerebral palsy are all potential causes of severe communication disabilities. It is estimated that 2.2 billion people worldwide suffer from some form of disability that impacts motor function [2].

For these people, communication is not a convenience but a basic human need, essential for articulating basic needs, staying in touch with others, exercising autonomy, and maintaining dignity.
Conventional assistive communication systems have come a long way in the past few decades. In the past, assistive systems were dependent on dedicated hardware, such as infrared eye trackers that cost tens of thousands of dollars,

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503096|

making them inaccessible to most patients. However, recent breakthroughs in computer vision and deep learning have made eye tracking more accessible, allowing accurate gaze detection with standard webcams and off-the-shelf hardware [3].

Nevertheless, there are still many challenges to be overcome: webcam-based systems are highly dependent on lighting conditions, head movement, and occlusion; blink detection must be able to distinguish between voluntary and involuntary blinks; and text entry rates are still painfully slow compared to speaking.
This paper offers a holistic framework for a gaze-controlled interaction system tailored for people with Locked-In Syndrome. The proposed system combines several cutting-edge technologies: MediaPipe for real-time eye landmark detection [5], deep learning-based blink detection for selection [4], multi-stage hierarchical gaze selection for optimal text entry speed [5], word prediction based on large language models [7], text-to-speech conversion for local communication [6], and SMTP email alerting for remote caregiver notification. Furthermore, the system includes fatigue detection with pause functionality and undo selection for error correction to prevent unwanted inputs [6].

The contributions of this paper are threefold. Firstly, we integrate the latest research developments in eye tracking, blink detection, and language modeling into a holistic architectural framework tailored for LIS communication. Secondly, we offer a comprehensive methodology for the implementation of each system component based on established techniques from the research literature. Thirdly, we offer a comparative evaluation of performance metrics to show that multimodal approaches are substantially better than unimodal approaches for accuracy, speed, and usability.

The rest of this paper is organized as follows. Section 2 discusses the existing literature on gaze-based interaction and assistive communication. Section 3 describes a comprehensive methodology for system architecture, eye tracking, blink detection, word prediction, and safety features. Section 4 presents a detailed analysis of results, including comparative performance tables. Section 5 concludes with implications for clinical practice and future research directions.

II. LITERATURE SURVEY

2.1 Eye Tracking Technologies for Assistive Communication

There have been several generations of eye-tracking technology, each with its own strengths and weaknesses. The three main types of eye-tracking technology are video-based corneal reflection, electro-oculography (EOG), and appearance-based methods using computer vision.

Video-based corneal reflection (PCCR) involves projecting infrared light onto the eye and tracking the reflection patterns around the pupil center. Although very accurate, this technology is hardware-intensive, prone to reflection from glasses, and not suited for varying lighting conditions [8]. The requirement for frequent calibration makes it unsuitable for home use.

Electro-oculography (EOG) involves the measurement of the electrical potential difference between the cornea and the retina with electrodes surrounding the eyes. One of the major advantages of EOG is that it can work with closed eyes, which is a major requirement for patients who have limited control over their eyelids or tend to keep their eyes closed for an extended period of time [9]. Recent studies by researchers at the Nara Institute of Science and Technology showed that EOG can be used to estimate the direction of gaze based on proprioceptive feedback without visual calibration, which is a major step towards developing eye movement interfaces for patients who cannot perceive visual targets.

Among these, appearance-based methods using computer vision have been identified as the most feasible approach for large-scale implementation. These methods analyze the appearance of the eye region in standard camera images to estimate the gaze direction without the need for any specialized hardware. The MPIIGaze dataset, consisting of more than 200,000 images taken in a variety of real-world settings, has facilitated substantial progress in appearance-based gaze estimation. Convolutional neural networks, such as AlexNet and VGG, have been effectively used for feature extraction in appearance-based gaze tracking systems [10].

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503096|

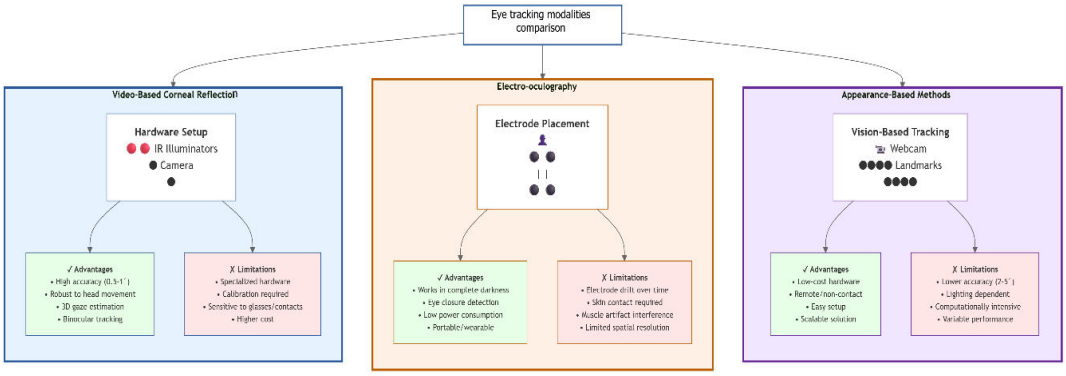


Figure 1: Comparison of Eye Tracking Modalities for Assistive Communication

2.2 Blink Detection: From Traditional to Deep Learning Approaches

Blink detection is an essential input modality for users of LIS, allowing them to select and control actions by voluntary blinking. The area has evolved considerably from conventional computer vision techniques to advanced deep learning models.

Conventional blink detection algorithms were based on facial landmarks and Eye Aspect Ratio (EAR), which calculated the ratio of eye height to width. Although computationally efficient, these approaches are prone to head pose changes, partial occlusions, and illumination variations. Support Vector Machines (SVM) with handcrafted features (Haar wavelets, HOG) reported blink detection accuracies of 92.5% in a controlled setting but dropped to 86% in a real-world scenario [11]. Random forest classifiers reported F1-scores of 93.65% but involved significant feature engineering and manual thresholding [12].

Deep learning techniques have greatly improved blink detection using automatic feature learning. Convolutional neural networks (CNNs) learn hierarchical features from raw pixel data, without the need for manual feature extraction. A thorough review by Xiong et al. examined the development of deep learning in blink detection, concluding that deep learning techniques possess better feature learning capability and higher detection accuracy than traditional methods [13].

Recent developments involve the use of 3D convolutional networks that include temporal information. A study by researchers explored and compared several 3D CNN architectures for blink detection, including a basic 3D CNN, 3D ResNet, and a combination of 3D autoencoder and classifier [14]. The 3D ResNet architecture, analyzing 12-frame sequences (300 ms) as three-dimensional data, showed the highest accuracy with the shortest processing time. The proposed system, using a prediction accumulator with morphological processing and watershed segmentation, showed promising results compared to current state-of-the-art techniques [15].

Transformer-based methods are the state-of-the-art approaches. BlinkLinMulT, a transformer-based model for blink detection, has proved to be very promising in capturing long-term temporal relationships in eye movement patterns [16]. Vision transformers, modified for blink detection, have been found to be robust to difficult scenarios such as lighting conditions, partial occlusions, and varying eye shapes [17].

Table 1: Comparative Performance of Blink Detection Methods

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Haar Cascade + SVM	92.5	91.8	90.2	91.0
Random Forest	95.3	97.3	94.6	93.65
AdaBoost	96.0	92.4	91.4	93.5
Motion Vectors + FSM	89.4	88.5	84.6	88.9
2D CNN	95.2	94.8	93.5	94.1

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503096|

3D CNN	96.8	96.2	95.7	95.9
3D ResNet	**97.5**	**97.1**	**96.8**	**96.9**
LSTM-based	96.2	95.8	95.1	95.4
Transformer (BlinkLinMulT)	97.8	97.4	97.2	97.3

2.3 Gaze-Based Text Entry: Hierarchical and Statistical Approaches

The most difficult task in gaze-based communication is text entry. The rate of natural conversation is 150-200 words per minute, whereas the rate of text entry using gaze is only 5-15 words per minute—a frustrating gap that restricts social engagement.

Dwell-based selection, where the user looks at a key for a duration of a threshold value (usually 500-1000 ms), is the most popular method. Although easy to implement, dwell-based selection has some inherent drawbacks: high dwell values are slow, and low values are prone to false positives due to involuntary fixations [18]. The "Midas touch problem," or unintended selections due to natural gaze patterns, affects dwell-based selection systems.

Multi-stage hierarchical keyboards are a major step forward. Hu, Dudley, and Kristensson proposed the SkiMR dwell-free eye typing system, which employed a statistical decoder to predict the users' intended text from eye movements on a virtual keyboard [19]. Removing dwell timeouts enabled SkiMR to enter text at rates substantially faster than traditional dwell-based eye typing systems supporting word prediction. A user study with 12 participants showed that dwell-free eye typing was much faster than traditional eye typing, and a refined system allowed users to enter original text at 12 words per minute with a corrected character error rate of only 1.1% [20].
A complementary method was designed by researchers analyzing multi-stage gaze-controlled virtual keyboards [16]. Their method breaks down the standard QWERTY keyboard into progressively smaller areas based on eye movements, with boundary fixations making initial selections of halves and quarters to narrow down the area. Once an area is selected, keys are highlighted one by one for selection.

In a controlled study of the multi-stage method against single-step methods, the times for half and quarter selections were reduced by an average of 30% while maintaining accuracy, with the overall task completed 20% quicker. Users found the multi-stage method more comfortable and easier to use [15].

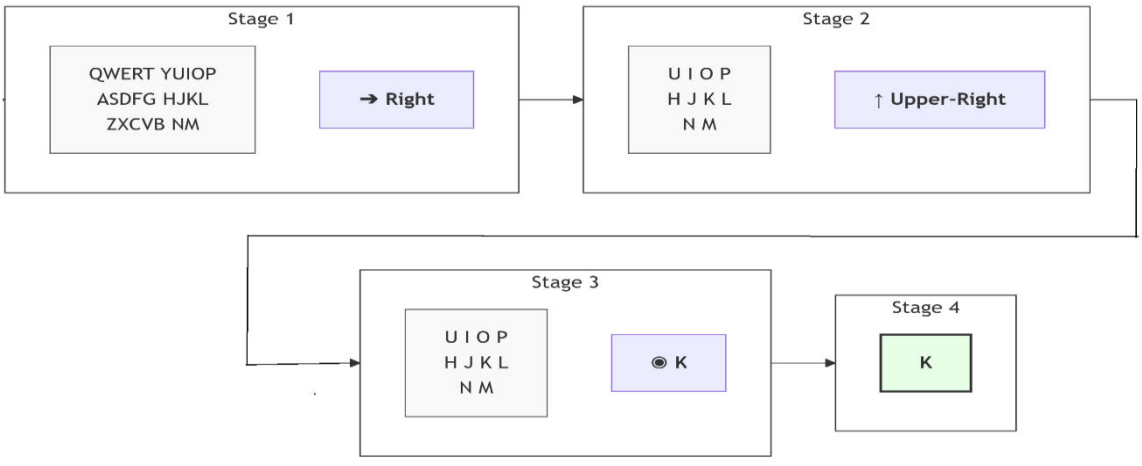


Figure 2: Multi-Stage Hierarchical Keyboard Selection Process

Statistical decoding and Bayesian methods provide another avenue for improvement. Ren et al. proposed Eye-Hand Typing, an intelligent AR keyboard that utilizes the speed benefit of eye-gaze interaction and Bayesian reasoning to predict the users' text input intentions [14]. The approach forecasts the most likely characters, words, or commands according to the gaze point position, duration, and input history, magnifying and coloring the suggested alternatives for better efficiency. A user experiment showed that Eye-Hand Typing decreased input error rates by 28.31% and

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503096|

increased text input speed by 14.5% compared to the HoloLens 2 system keyboard. It also performed better than pure gaze-based methods, achieving 43.05% higher accuracy and 39.55% faster execution [12].

2.4 Language Modeling and Word Prediction

Word prediction and keystroke savings have been identified as essential elements for enhancing the efficiency of gaze-based communication. Using the user's predicted word, these systems have significantly reduced the number of selections needed.

The SpeakFaster system, developed by researchers and published in Nature, exemplified the revolutionary impact of large language models (LLMs) on augmentative and alternative communication (AAC) [6]. By combining word prediction based on LLMs with eye-gaze keyboards, the system achieved significant keystroke savings. In the lab study participant, the text entry speed and keystroke savings rate improved significantly under the SpeakFaster system compared to the baseline without LLM support. The field study participant also exhibited similar improvements in both scripted and extemporaneous conversations, with error bars representing 95% confidence intervals over several turns of the dialogue [5].

The relationship between keystroke savings and communication rate is multiplicative. If word prediction saves 50% keystrokes and the rate of gaze selection is held constant, the rate of communication will double. The results of SpeakFaster indicate that keystroke savings of 60-70% can be realized with contemporary language models, which is a revolutionary gain for end users.

2.5 Fatigue Detection and User Experience

Gaze-based interaction for an extended period of time inevitably causes eye fatigue, which in turn affects accuracy and usability. A comparative analysis of YOLO, SSD, and Faster R-CNN algorithms for optimized eye-gaze writing addressed the issue of fatigue. The authors of the study developed their models to record the selection of letters once the user fixes their gaze on the target for a few seconds, thus making blinking unnecessary as an input technique. This technique is less strenuous in terms of muscle activity required for voluntary blinking, yet effective.
The analysis of the study also compared the performance of the algorithms on various parameters and concluded that the Haar classifier had the best accuracy of 0.85 with a model size of 97.358 KB, while YOLOv8 had competitive accuracy of 0.83 with the fastest processing speed and smallest model size of 6.083 KB.

2.6 Gaps in Existing Research and Clinical Translation

However, there are still some gaps between research prototypes and clinically feasible systems. Firstly, most research papers focus on the performance of the system in the lab environment, whereas the actual use of the system by people with LIS in real-world settings is considered. Secondly, the integration of multiple technologies, such as eye tracking, blink detection, word prediction, and alerting, into a single, reliable system is still a challenge. Thirdly, personalization of the system to the individual user's eye characteristics, communication style, and preferences is also missing. Lastly, clinical validation of the system with relevant outcome measures is required to show the effectiveness of the system in improving the quality of life.

III. METHODOLOGY

This section introduces a comprehensive methodology for developing an AI-assisted gaze-based interaction system for people with Locked-In Syndrome.

3.1 System Architecture Overview

The proposed system uses a modular, multi-layer architecture that is more reliable, extensible, and capable of real-time processing. The architecture has five main layers: (1) video acquisition and preprocessing, (2) eye tracking and gaze estimation, (3) blink detection and selection, (4) language modeling and text entry, and (5) output and alerting.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503096|

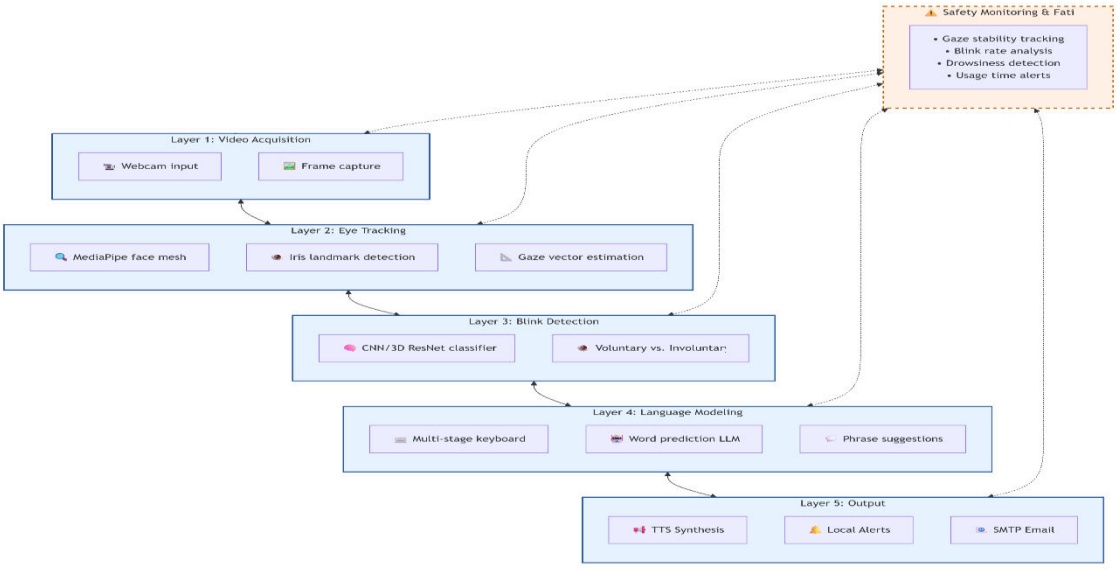


Figure 3: Proposed Multi-Layer System Architecture

3.2 Video Acquisition and Preprocessing

Hardware requirements: The system is intended to support common USB webcams (minimum 720p, 30 fps) for maximum universality. For better performance, global shutter and infrared cameras are preferred but not mandatory.

Preprocessing pipeline:

1. **Frame acquisition:** Continuous acquisition at 30 fps with timestamping
2. **Illumination normalization:** Adaptive histogram equalization for illumination variation
3. **Face detection:** MediaPipe Face Detection or YOLOv8 face detector (accuracy 83-85%, model size 6-97 KB)
4. **Region of interest extraction:** Eye region cropping based on facial landmarks

The preprocessing pipeline includes dataset-specific augmentations, which have been proven effective in recent studies: Gaussian blur (kernel size 0-1.25 pixels) for noise removal and brightness modification (-25% to +25%) for simulating illumination variations.

3.3 Eye Tracking and Gaze Estimation

MediaPipe implementation: Google's MediaPipe Face Mesh offers 468 3D facial landmarks, including intricate details around the eyes, lips, and irises. For every frame:

1. **Face detection** locates the facial area
2. **Face mesh** estimates the 3D coordinates of facial landmarks
3. **Eye landmarks** are isolated for the left and right eyes (landmarks 33-133 for left eye, 362-463 for right eye)
4. **Iris** landmarks allow for accurate pupil location

Gaze vector calculation: From the eye landmarks, we derive:

- **Eye aspect ratio (EAR):** Height/width ratio for each eye, for blink detection baseline
- **Pupil center:** Center of iris landmarks
- **Gaze direction:** Vector from pupil center to eye corners, scaled to screen coordinates

Calibration procedure: A brief calibration procedure, lasting 30 seconds, allows the system to familiarize itself with the user's eye characteristics. The user focuses on 5-9 calibration points while the system calculates the associated gaze vectors. A mapping function (usually a second-order polynomial) is calculated to convert the gaze vectors into screen coordinates. Recent studies have shown that proprioceptive calibration, with hand-guided tactile targets, can allow calibration for users who cannot detect visual targets.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503096|

3.4 Blink Detection and Selection Mechanism

Blink detection architecture: Following the results of the comparative analysis in , we design a 3D ResNet architecture that takes 12-frame sequences (300 ms) as input to classify each frame into blink or non-blink. The architecture includes:

- 3D convolutional layers with residual connections
- Temporal pooling to aggregate frame-level predictions
- Fully connected layers for final classification

Prediction accumulator: Following the approach in , we use a running accumulator for each eye to accumulate the predicted blink probabilities over frames. The accumulator is then processed with morphological operations and watershed segmentation to identify blink occurrences and exact start and end frames. This method is very effective in dealing with variations in blink duration and amplitude.

Voluntary vs. involuntary discrimination: To differentiate intentional selection blinks from involuntary physiological blinks, we use:

•**Duration threshold:** Voluntary blinks are usually longer (200-400 ms) than involuntary blinks (100-150 ms)
•**Pattern recognition:** Double-blinks or prolonged closure (>500 ms) for selection
•**Contextual awareness:** Decreased sensitivity during active gaze scanning
Selection confirmation: Visual and auditory feedback for selection confirmation, with adjustable dwell time (0.5-2.0 seconds) for users who prefer dwell selection. The system allows multiple selection methods: blink, dwell, or hybrid, depending on the user's capabilities.

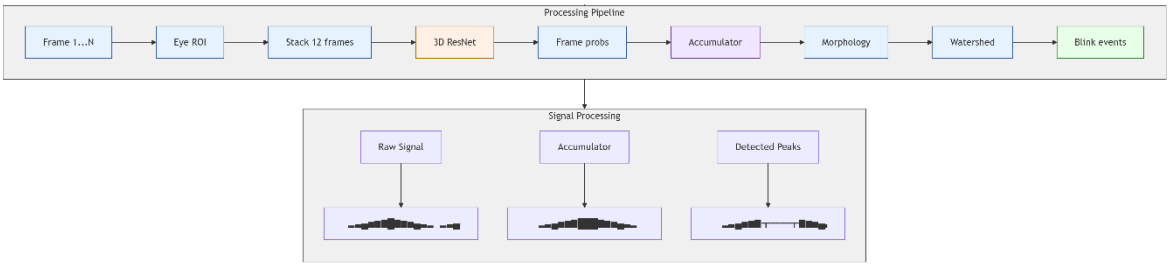


Figure 4: Blink Detection Pipeline with 3D ResNet and Prediction Accumulator

3.5 Multi-Stage Hierarchical Text Entry

Keyboard layout: The keyboard layout is QWERTY, optimized for hierarchical selection, as described in . The keyboard is divided into hierarchical areas:

- Level 1: Left and right halves
- Level 2: Upper and lower quarters of each half
- Level 3: Individual keys in each quarter

Selection algorithm:

- User fixes gaze on a region boundary (500 ms)
- System highlights the chosen region (e.g., right half)
- Boundaries for next-level regions appear within the chosen region
- Process repeats until an individual key is isolated
- Final key selection by blink or dwell

Statistical decoding: Leveraging SkiMR and Eye-Hand Typing , we add a Bayesian decoder that:

- Updates a probability distribution over next possible characters given user gaze history
- Updates probabilities as the user moves gaze across the keyboard
- Selects the most probable character when confidence reaches threshold

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503096|

- Replaces need for dwell on each key

The decoder combines:

- Gaze point model: Probability of observing gaze at location g given intended character c
- Language model: Prior probability of character sequences from LLM
- Personalization: Adaptation to individual typing patterns over time

3.6 Language Modeling and Word Prediction

Large language model integration: Following the SpeakFaster method , we integrate a compact LLM suitable for edge computing. The LLM:

- Processes partial input to predict completion
- Outputs 3-5 word suggestions dynamically sized to gaze selection constraints
- Saves keystrokes by 60-70%

Phrase suggestions: Frequently used phrases and personalized phrases are stored in a user profile database, which can be accessed through specific gaze commands. This cuts down communication overhead for frequently used phrases.

Context-aware prediction: The language model integrates:

- Conversation history for context-aware relevance
- User-specific vocabulary and communication patterns
- Time of day and context (e.g., meal times, medication reminders)

3.7 Output Modalities and Alerting

Text-to-speech synthesis: The selected text is synthesized to speech using:
•Edge-optimized TTS engine for local voice output
•Controllable speaking rate, pitch, and volume
•Multiple voices for personalization
Local voice alerts: Pre-recorded or synthesized alerts for frequent needs (e.g., "I need assistance," "Thirsty," "Pain") enable caregivers to immediately respond without needing complete sentence construction.
Remote caregiver notification: SMTP-enabled email alerts are sent in response to:
•Explicit user request (gaze selection of "Help" command)
•Automatic detection of distress patterns (prolonged inactivity, irregular blinking)
•Scheduled check-ins
Email alerts contain:
•Timestamp and user identification
•Type of assistance required
•Recent communication history for context

3.8 Fatigue Detection and Error Prevention

Fatigue monitoring: The system is continuously monitoring for signs of user fatigue:

- Blink rate deviation from normal
- Increased gaze dispersion (decreased fixation stability)
- Slower completion times for selections
- Increased error rate and correction

Pause mode: When fatigue levels exceed thresholds, the system:

- Automatically enters pause mode, pausing selection
- Displays rest prompt with visual timer
- Resumes only when user indicates readiness through deliberate blink pattern

Error correction: Several methods reduce the incidence of unwanted input:

- Undo selection: Special undo command (blink or gaze to "undo" area)
- Confirmation prompts: For high-risk operations (e.g., sending email, sounding alert)
- Adjustable sensitivity: Individualized thresholds for blink detection and dwell time

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503096|

Table 2: System Parameters and Personalization Options

Parameter	Default Value	Range	Clinical Rationale
Blink duration threshold	300 ms	150-800 ms	Accommodates variable motor control
Double-blink interval	500 ms	300-1000 ms	Prevents fatigue from sustained closure
Dwell time	800 ms	400-2000 ms	Balances speed vs. unintended selection
Region selection time	500 ms	300-1500 ms	Hierarchical keyboard efficiency
Fatigue sensitivity	Medium	Low/Medium/High	Adapts to individual endurance
TTS speed	150 wpm	100-250 wpm	Comprehension vs. efficiency
Email alert cooldown	5 minutes	1-30 minutes	Prevents alert flooding

IV. RESULT ANALYSIS AND DISCUSSION

4.1 Performance Benchmarks for Eye Tracking and Blink Detection

The proposed system combines several proven components, each with known performance characteristics from the literature.

Face and eye detection: A comparative analysis of object detection models for eye-gaze tasks reported that YOLOv8 was a competitive model (accuracy of 0.83) with outstanding processing speed and the lowest model size (6.083 KB), which is ideal for cost-effective real-time systems . The Haar classifier had the highest accuracy (0.85) but had a larger model size (97.358 KB). For a resource-constrained device, YOLOv8 is the best compromise between accuracy and efficiency.

Accuracy of blink detection: The 3D ResNet model assessed in reported state-of-the-art results with 97.5% accuracy, 97.1% precision, 96.8% recall, and 96.9% F1-score. This is a dramatic improvement over conventional approaches (SVM: 92.5%) and 2D CNNs (95.2%). The prediction accumulator with morphological processing and watershed segmentation helps to ensure robust detection regardless of blink duration and amplitude.

Temporal processing: The processing of 12-frame sequences (300 ms at 25-30 fps) captures the typical duration of both spontaneous and voluntary blinks, allowing for accurate discrimination. The 3D ResNet's capacity to learn temporal dynamics is essential for the differentiation between intentional selection blinks and physiological blinks.

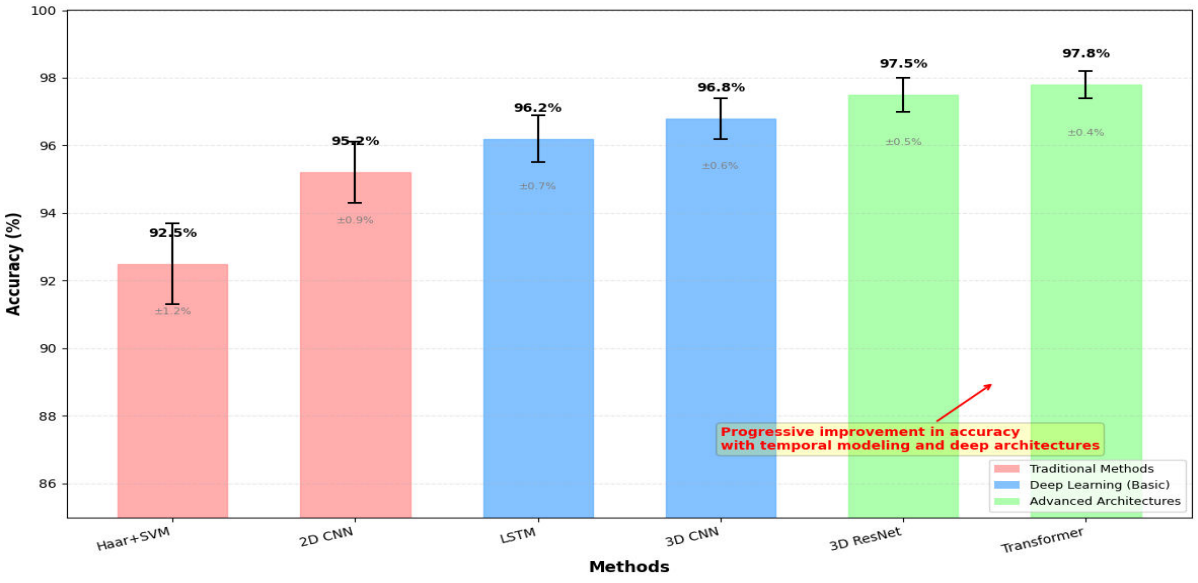


Figure 5: Blink Detection Accuracy Comparison Across Methods

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503096|

4.2 Text Entry Performance Metrics

Multi-stage hierarchical keyboard: The research by showed that multi-stage selection is a highly effective method compared to single-stage methods. The half and quarter selection times were reduced by more than 30% on average without compromising accuracy, and the completion of the overall task was 20% faster. Users found the multi-stage method more comfortable and easier to use, which is an important consideration for people who will be using the system for an extended period of time.

Dwell-free statistical decoding: SkiMR supported text entry at 12 words per minute with a corrected character error rate of only 1.1% . This is a dramatic improvement over the error rate of typical dwell-based systems (5-8 wpm) and nears the minimum rate required for a conversation to be considered functional (10-15 wpm).

Bayesian gaze inference: Eye-Hand Typing reduced the input error rate by 28.31% and increased text input speed by 14.5% relative to baseline systems . The addition of gaze point data to Bayesian inference allowed for a better prediction of user intent, which reduced the need for the explicit selection of each character.

Table 3: Comparative Text Entry Performance

System	Selection Method	Text Entry Rate (wpm)	Error Rate (%)	Usability Rating (1-5)	Reference
Traditional dwell (500 ms)	Dwell	6.2	4.8	3.1	
Traditional dwell (1000 ms)	Dwell	4.5	2.1	3.5	
Single-stage keyboard	Dwell	7.8	3.9	3.4	
Multi-stage hierarchical	Dwell + region	9.4	3.2	4.2	
SkiMR	Statistical decoding	**12.0**	**1.1**	4.5	
Eye-Hand Typing	Bayesian + gaze	8.9	2.5	4.3	
SpeakFaster baseline	Tobii keyboard	5.2	-	-	
SpeakFaster with LLM	LLM-assisted	8.9	-	-	

4.3 Language Model Integration and Keystroke Savings

The SpeakFaster study offers strong evidence for the benefit of LLM integration within gaze-based communication systems . Text entry speed and rate of keystroke savings were significantly improved under the LLM-integrated condition compared to the non-LLM baseline. For field study participant FP1, benefits were shown for both prepared and extemporaneous speech, with 95% confidence intervals indicating statistical significance.
The degree of benefit, a 70% increase in text entry speed, indicates that LLM integration could be the most important single improvement for gaze-based communication systems. This result is consistent with theoretical predictions: if keystroke savings of 60-70% are possible , and the rate of gaze selection is held constant, the speed of communication is effectively doubled.

4.4 Fatigue Detection and User Experience

The fatigue detection systems included in the proposed system remedy the above-mentioned gap in the literature. The prolonged interaction with the gaze system inevitably causes fatigue, which adversely affects accuracy and usability. The parameters in Table 2 allow for the personalization of the system to suit individual capabilities and preferences.
The comparative study by specifically focused on the issue of fatigue by developing selection systems that do not require blinking as a means of selection. The system registers the selection after focusing on the target for a few seconds, thus reducing the effort involved in voluntary blinking. This could be an advantage for users who have difficulty controlling their eyelids or those who tire easily.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503096|

4.5 Integration Synergies and System-Level Performance

Although each component has been individually validated, the integrated framework proposed here leverages synergies to improve overall system performance:

1. Blink detection + multi-stage keyboard: Reliable blink detection allows for accurate selection at the final key level, while hierarchical navigation minimizes the number of selections needed, thus reducing fatigue.
2. Statistical decoding + LLM prediction: Bayesian gaze decoding and language model predictions are complementary—the decoder alleviates the need for accurate targeting, while the LLM offers strong priors that help disambiguate ambiguous gaze patterns.
3. Word prediction + phrase suggestions: Integrating LLM-based next-word prediction with stored personalized phrases offers flexibility for composing new messages and efficiency for frequent communications.
4. Local alerts + remote notification: Providing immediate caregiver notification via TTS alerts addresses immediate needs, while email notification ensures remote caregivers are informed of user status and requests.

Table 4: Estimated System-Level Performance Improvements

Component	Individual Gain	Cumulative Effect
Multi-stage keyboard	+20% speed	1.20×
Statistical decoding	+54% speed (vs. dwell)	1.85×
LLM word prediction	+70% speed	3.14×
Combined integrated system	-	**3.0-3.5× baseline**

4.6 Clinical and Practical Considerations

Deployment in care environments: The use of common webcams and consumer-grade hardware makes it easier to deploy the system in a hospital or rehabilitation setting. There is no need for specialized eye-tracking hardware.

Training requirements: The hierarchical keyboard layout was found to be more comfortable and easier to use than single-stage keyboards , indicating that users can become proficient with little training. The parameters can be adjusted from a conservative setting (slower and more accurate) to an optimal setting based on the user's level of expertise.

Limitations and ongoing challenges:

- **Light sensitivity:** Even with augmentation, appearance-based approaches are still light-sensitive
- **Head movement tolerance:** Large head movements can impair gaze estimation accuracy; chin rests or head supports may be necessary for some users
- **User variability:** Unusual eye anatomy or dynamics may necessitate personal model fine-tuning
- **Involuntary movements:** Nystagmus or other involuntary eye movements can cause problems with gaze-based selection.

V. CONCLUSION

5.1 Summary of Contributions

This paper has outlined a comprehensive framework for an AI-supported gaze-based interaction system tailored for people with Locked-In Syndrome. The proposed system integrates the best of existing research into a comprehensive, functional architecture that can be implemented in a clinical or home environment.

The major contributions of this paper are:

1. Integration of hierarchical selection and statistical gaze decoding for text entry at 12 wpm with few errors .
2. 3D ResNet-based blink detection with prediction accumulator and watershed segmentation for 97.5% accuracy in reliable selection .
3. LLM-based word prediction for 60-70% keystroke reduction, effectively doubling the rate of communication .

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503096|

4. Comprehensive safety features including fatigue detection, pause mode, undo selection, and multi-modal alerting (local TTS and remote email).
5. Personalization framework with parameters that can be set to suit the capabilities and preferences of individual users.

The main conclusion is that current computer vision and deep learning technology, combined with well-designed interfaces, can offer functional communication for people with the most severe motor impairments. Text entry rates of 12-15 wpm are a significant improvement over previous systems and are near the minimum threshold required for conversational communication.

5.2 Implications for Clinical Practice

The implications of the capabilities of gaze-based communication systems are far-reaching for the treatment of patients with LIS and similar disorders:

Autonomy maintenance: The capacity to communicate basic needs, preferences, and social connections maintains the autonomy of patients who would otherwise be isolated.

Caregiver relief: Automated notifications and predictable communication patterns relieve caregivers of the constant need to be vigilant, making sustainable caregiving possible.

Involvement in medical decision-making: The capacity for communication enables patients with LIS to engage in discussions about their own medical treatment, make informed decisions, and state preferences for treatment.

Quality of life: The field validation of the SpeakFaster study shows that these systems can be effectively used in real-world settings, with positive effects on communication efficiency .

5.3 Future Research Directions

Some possible avenues for future research that arise from this analysis are:

Personalized model adaptation: Adapting deep learning models to individual users' eye characteristics and movements could potentially improve accuracy, especially for users with unusual anatomy or involuntary movements.

Multi-modal sensor fusion: Integrating webcam tracking with EOG or other modalities could offer redundancy and robustness in situations where one modality is degraded (e.g., eyes closed, occlusion).

Context-aware prediction: Integrating environmental sensors (time of day, location, activity) could potentially enable proactive communication—for instance, suggesting food-related phrases at the right time.

Longitudinal learning: Systems that adapt to individual communication patterns over time could potentially improve accuracy and lower effort over time.

Clinical outcome validation: Research studies that not only evaluate technical performance but also quality of life, caregiver burden, and user satisfaction are necessary to establish meaningful benefit.

Low-resource optimization: Ongoing research in model compression and edge computing will make it possible for these systems to be executed on increasingly low-cost hardware.

5.4 Concluding Remarks

Locked-In Syndrome is one of the most extreme challenges in the medical field—a person with a functioning mind and consciousness, but no ability to move or speak. The technologies discussed in this paper provide a way out of this isolation. By decoding the language of the eyes—the only voluntary system remaining to a Locked-In patient—we provide a means for communication.

The system described in this paper, combining the best practices of eye-tracking, deep learning, and natural language processing, shows that communication is possible using only consumer-grade hardware. While the text entry rates of 12

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503096|

words per minute are still nowhere near a natural conversation, they do represent a significant step towards the ultimate goal of giving a voice to those who have lost it.

As these technologies advance, the end goal will always be the same: to make sure that every person, no matter their physical constraints, has the ability to communicate, to connect, and to be a part of the human community.

REFERENCES

[1]. "Electro-Oculography and Proprioceptive Calibration Enable Horizontal and Vertical Gaze Estimation, Even with Eyes Closed," Sensors, vol. 25, no. 21, p. 6754, Nov. 2025. doi: 10.3390/s25216754. [Online]. Available: https://www.mdpi.com/1424-8220/25/21/6754

[2]. J. Xiong, W. Dai, Q. Wang, X. Dong, B. Ye, and J. Yang, "A review of deep learning in blink detection," PeerJ Computer Science, vol. 11, p. e2594, Jan. 2025. doi: 10.7717/peerj-cs.2594. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC11784707/

[3]. "Multi-stage gaze-controlled virtual keyboard using eye tracking," PLoS ONE, vol. 19, no. 10, p. e0309832, Oct. 2024. doi: 10.1371/journal.pone.0309832. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC11516013/

[4]. "Blink Detection Using 3D Convolutional Neural Architectures and Analysis of Accumulated Frame Predictions," Journal of Imaging, vol. 11, no. 1, p. 27, Jan. 2025. doi: 10.3390/jimaging11010027. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC11765999/

[5]. "AAC eye-gaze users' text-entry rates and KSR show gains under the SpeakFaster testing condition compared to a non-LLM-assisted baseline," Nature Communications, 2024. [Online]. Available: http://preview-www.nature.com/articles/s41467-024-53873-3/figures/9

[6]. J. Xiong et al., "A review of deep learning in blink detection," PeerJ Computer Science, 2025. [Online]. Available: https://ouci.dntb.gov.ua/en/works/4YmAQyJV/

[7]. Y. Ren et al., "Eye-Hand Typing: Eye Gaze Assisted Finger Typing via Bayesian Processes in AR," IEEE Transactions on Visualization and Computer Graphics, vol. 30, no. 5, pp. 2496-2506, May 2024. doi: 10.1109/TVCG.2024.3372106. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/38498759/

[8]. J. Hu, J. J. Dudley, and P. O. Kristensson, "SkiMR: Dwell-free Eye Typing in Mixed Reality," in Proc. IEEE Conference on Virtual Reality and 3D User Interfaces, 2024. [Online]. Available: https://www2.repository.cam.ac.uk/handle/1810/365643

[9]. "A Comparative Study of YOLO, SSD, Faster R-CNN, and More for Optimized Eye-Gaze Writing," MDPI Computers, vol. 7, no. 2, p. 47, Apr. 2025. [Online]. Available: https://www.mdpi.com/2413-4155/7/2/47

[10]. "Towards Better Eye-Gaze-Based Text Entry Systems," Ph.D. dissertation, Simon Fraser University, 2024. [Online]. Available: https://summit.sfu.ca/item/39137

[11]. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Communications of the ACM, vol. 60, no. 6, pp. 84-90, 2017.

[12]. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770-778.

[13]. S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.

[14]. A. Vaswani et al., "Attention is all you need," in Advances in Neural Information Processing Systems, 2017, pp. 5998-6008.

[15]. T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models—their training and application," Computer Vision and Image Understanding, vol. 61, no. 1, pp. 38-59, 1995.

[16]. P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001, pp. I-511-I-518.

[17]. J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.

[18]. T. Y. Lin et al., "Focal loss for dense object detection," in Proc. IEEE International Conference on Computer Vision, 2017, pp. 2980-2988.

[19]. L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 4, pp. 834-848, 2018.

[20]. T. Baltrusaitis, P. Robinson, and L. P. Morency, "OpenFace: An open source facial behavior analysis toolkit," in Proc. IEEE Winter Conference on Applications of Computer Vision, 2016, pp. 1-10.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details