**Enhancing robotic surgery through eye gaze-controlled camera navigation for reduced cognitive load**

**Balabadruni Venkata Naga koushik1, Saranyaraj D1, Gade Sai Laxmi Ganesh Reddy1, Potturi Pujeth1**

1School of computer science and engineering, Vellore Institute of Technology, Chennai, India.

Corresponding author: Saranyaraj .D ([saranyaraj.d@gmail.com](mailto:saranyaraj.d@gmail.com));

Contributing authors: Balabadruni Koushik ([Venkata.nagakoushik2021@vitstudent.ac.in](mailto:Venkata.nagakoushik2021@vitstudent.ac.in)); Saranyaraj D ([saranyaraj.d@vit.ac.in](mailto:saranyaraj.d@vit.ac.in)); Potturi Pujeth ([potturi.pujeth2021@vitstudent.ac.in](mailto:potturi.pujeth2021@vitstudent.ac.in)); Gade Ganesh Reddy([sai.laxmiganesh2021@vitstudent.ac.in](mailto:sai.laxmiganesh2021@vitstudent.ac.in)).

Abstract

This study investigates the potential of eye-gaze detection to enhance dynamic camera control in robotic surgeries, aiming to decrease cognitive load on surgeons during challenging procedures in the realm of robotic surgery. Standard robotic surgical systems, such as Da Vinci and ZEUS, require the surgeon not only to operate the robotic arms but also to control the camera, which requires mental focus and splits the surgeon's attention. This split attention can lead to cognitive overload in complex operations. Therefore, the paper suggests a solution providing an autonomous camera control, which takes the camera position according to the eye-gaze of the respective surgeon, thereby facilitating the work of the surgical actions. The setup consists of a Raspberry Pi 3 and a custom-trained YOLOv9 model to detect the direction of gaze of the surgeon in real time: left, right, up, down, and center. A thorough investigation and training of the YOLOv9 model were carried out for 100 epochs, and the system was found to be highly accurate, achieving 90.78% accuracy (mAP50 in context of YOLO) in the inference. Alongside this training, comparative studies regarding various models of YOLO, such as YOLOv11 and earlier versions, were conducted to find a suitable model for the proposed application. The results of that show the clear dominance of YOLOv9 compared to others, especially for 100 epochs, proving to be the best model for real-time eye-gaze tracking with limited resources. With the ability to provide accurate and dynamic camera adjustments according to eye movement, the system can greatly improve the efficiency of robotic-assisted surgeries by decreasing cognitive load on surgeons, thus allowing better focus on the task in hand. The study results emphasize the potential of integrating advanced machine-learning models like YOLOv9 for real-time gaze tracking providing the pathway to enabling a more efficient hands-free robotic surgery system with improvements to precision and experience for the surgeon.

Keywords: YOLO, Eye gaze, Raspberry pi3, Robotic surgeries.

1. Introduction

This breakthrough in eye-tracking technology has newly opened the gateway to new avenues of human-computer interaction, especially in fields wherein precision and hands-free control are crucial. One such application is the use of eye-gaze controllers for cameras and robotic systems. The research presents an eye-gaze detection with application using the YOLO object model in combination with a Gaze Tracking library [1]. It guarantees hands-free control of the camera. Despite advancements, a research gap exists in integrating robust, real-time gaze tracking for surgical robotics, which this work addresses. This study looks to propose a system where the movement of the user's eyes is monitored and translated in real-time into corresponding movements of the gimbal system. This setup avoids all the tedium of servos running cameras but gives far more intuitive and effective management in situations such as surgery, where it is necessary to track the field of view of a surgeon.

As robotic-assisted surgeries evolve, studies predict a 40% increase in their adoption due to improvements in AI-driven automation and precision [39]. The integration of eye-gaze tracking systems, as proposed in this study, aligns with these advancements by reducing cognitive load on surgeons. The Da Vinci single-port robotic system has demonstrated a 21.4% reduction in incision size and a 17.6% decrease in recovery time [40]. Our gaze-controlled camera adjustment system aims to complement such innovations by enhancing surgical precision and workflow efficiency.

The YOLO model features a high speed and accuracy in the task of object detection within an image; therefore, it is used for real-time eye-position detection from the user. The YOLO model, having been specifically trained for this task, enables the system to quickly identify the gaze location and direction with minimal latency between the eye movement of the user and the response of the camera [2]. In contrast to other detectors' models, since YOLO processes frames at a relatively high speed, it is appropriate for dynamic environments in which the attention of a user can change very rapidly.

Gaze tracking accuracy is improved by adopting the library Gaze Tracking [3] into the developed system. It provides a very simple yet efficient technique to detect and interpret the presence of gaze patterns using facial landmarks [4]. Combining this information with eye position obtained through YOLO and gaze data obtained from GazeTracking, the system was able to correctly determine the focus of attention and accordingly actuate the gimbal. The synergy of these two technologies provides an efficient solution for moving the camera because of changes in lighting and movement conditions [5].

Positioning a camera in robotic-assisted surgery is therefore important in ensuring the surgeons an unobstructed, high-resolution view of the operative field. Current conventional methods of solving camera placement issues result in manual adjustment of the camera, which disrupts the fluidity of the surgery and introduces cognitive load as the surgeon has to shift focus between the surgical task and camera control. This eye-gaze-based system offers the solution as it lets the camera trace the gaze of the surgeon immediately, thereby taking away the need for manual adjustments [6]. Such hands-free control could potentially make room to streamline workflows in the operating room to let surgeons focus on the procedure at times when the competition in terms of time is zero for it to occur.

YOLO's real-time performance efficiency makes the application of this system in eye-gaze detection valuable [7]. YOLO stands for "You Only Look Once" and is meant for real-time object detection where any delays might result in an adverse impact on functionalities. Minor delays in moving cameras in surgical environments may also affect accuracy and effectiveness of the procedure [8]. It could then benefit maximally from YOLO's speed to remain in pace with the human eye movements, and this would make the experience even more fluid and sensitive, therefore further heightening the surgeon's control over the camera [9].

The GazeTracking library complements this precision capability with the ability to track specific facial landmarks [10]. This would allow the system to make subtle changes in interpreting the direction of the gaze and to adjust based on the corresponding change and, hence increase reliability even if variations in lighting conditions and the position of the user exist, such as in the case of fluctuating lighting in an operating room or a surgeon's posture shifting while operating. These are typical challenges to a system for tracking gaze. However, this would always maintain constant performance due to the confluence of strength in YOLO with the accuracy of GazeTracking. This multi-layered element also ensures that the camera follows the surgeon's gaze correctly and, therefore less likely to become misaligned.

Recent advancements in eye-gaze tracking for robotic surgery have demonstrated the potential of gaze-based systems to improve surgical efficiency and reduce cognitive load. For instance, Kwok et al. (2012) introduced Collaborative Gaze Channelling (CGC), which uses eye gaze to enhance cooperation in robotic-assisted surgery, showing significant improvements in surgical speed and accuracy [30]. Similarly, Ezzat et al. (2021) developed a gaze-based robotic scrub nurse (RSN) that allows surgeons to interact hands-free, reducing frustration and improving workflow efficiency [31]. These studies highlight the importance of integrating gaze-tracking technologies into surgical systems to enhance precision and reduce the cognitive burden on surgeons.

Moreover, Pan et al. (2024) proposed a deep learning-based eye-gaze control system for a needle deployment robot, achieving high precision in real-time eye fixation tracking [32]. This approach eliminates the need for preoperative training and intraoperative sterilization, making it highly suitable for minimally invasive surgeries. The success of such systems underscores the potential of combining deep learning models like YOLOv9 with gaze-tracking technologies to create more intuitive and efficient surgical interfaces.

Outside of these surgical applications, this technology has broad implications outside of surgery, where hands-free control is required. Industrial automation operators could view cameras without manual input. They would be free to observe machinery or processes, which require continuous observation for their operation to improve continually. For example, it could be used in augmented and virtual reality where it employs eye-gaze tracking to provide an immersed environment whereby the system automatically responds to the focus of the user's gaze. The proposed system is an extension that can be used to assist people mobility-impaired in the sense of them having a means of interfacing with devices made from robotics using merely their gazes [11].

Even though YOLO has been proven to be a great tool for object recognition of diverse objects in an image [12], it seems to open fresh opportunities for applicability in real-time human-machine interfaces for tracking of person's gaze. The work presented here demonstrates the flexibility offered by YOLO beyond its standard use cases and presents its usage as a complete asset in all applications that require speed and accuracy. With the ease with which the Gaze Tracking library integrates with YOLO, this system can be made adaptable to any type of real-time tracking without significant complexity in hardware.

Some technical issues that arose in the course of development include compatible setup on Arduino and the YOLO model for execution on a Raspberry Pi [13]. A preliminary test was done on Arduino [14] because it was more practical; though it showed bottlenecks in performance, since the pathway for power and signal between the power supply and Arduino does cause bottlenecks. This resulted in fewer power and signal channels required; thus, communication with the camera control unit became much more fluid. Great improvements in response and accuracy were reaped from the shift in hardware architecture, which further reviews the selection of hardware in the development of real-time systems.

1. **Related study**

Zhu et al. (2024) presented a novel frequency-encoded eye-tracking smart contact lens. The lens uses a unique frequency-encoding method to track eye movements, which the authors claim offers advantages in terms of accuracy, comfort, and unobtrusiveness compared to traditional camera-based eye trackers. The research demonstrates the potential of smart contact lenses for advanced human-machine interaction [1].

YOLOv1 to YOLOv10 (n.d.): This arXiv preprint likely provides a comprehensive overview of the evolution of YOLO (You Only Look Once) object detection architectures, from the original YOLOv1 to the latest versions (up to v10 as the title suggests). It likely compares the different versions in terms of speed, accuracy, and other performance metrics, highlighting the advancements and trade-offs made in each iteration [2].

Toabanda et al. (2023) offers a survey of gaze tracking, covering different devices, libraries, and applications. It likely categorizes various eye-tracking technologies, discusses their strengths and weaknesses, and explores the diverse range of applications where gaze tracking is used, from human-computer interaction to medical diagnostics [3].

Soukupová et al. (2016) focused specifically on real-time eye blink detection using facial landmarks. It likely details a method for accurately and efficiently detecting eye blinks in video streams, which is a crucial component of many eye-tracking systems and applications [4].

Rakhmatulin & Duchowski (2020) explored the use of deep neural networks for low-cost eye tracking. It likely investigates how deep learning models can be trained to estimate gaze direction using readily available hardware like webcams, making eye tracking more accessible and affordable [5].

Zhang et al. (2024) described a real-time camera-based gaze-tracking system with dual interactive modes. The authors likely demonstrate how their system can be used in gaming applications, showcasing the potential of eye tracking to enhance user experience and provide new forms of interaction [6].

Ultralytics YOLOv9 Docs, this is the official documentation for YOLOv5. It provides detailed information about the YOLOv9 architecture, including its different versions, training procedures, and usage instructions. It's an essential resource for anyone working with YOLOv9 [7].

Kenawy et al. (2022) investigated preventable operating room delays in robotic-assisted thoracic surgery. While not directly related to eye tracking or YOLO, it highlights the importance of efficiency in surgical procedures and the potential role of technology in improving surgical outcomes [8].

YOLO Architectures Review (n.d.) Similar to [2], this arXiv preprint likely provided a comprehensive review of YOLO architectures, possibly focusing on the evolution and comparison of different versions up to YOLOv8 and YOLO-NAS [9].

Ye et al. (2023) proposed a low-cost, geometry-based eye gaze detection method using facial landmarks generated through deep learning. It likely details the algorithms and techniques used to estimate gaze direction from facial landmarks, offering a computationally efficient approach [10].

Moreno-Arjonilla et al. (2024) focuses on survey on eye tracking in virtual reality. It likely discusses the challenges and opportunities of integrating eye tracking into VR environments, exploring how it can enhance user immersion, interaction, and research possibilities [11].

YOLO Object Detection Guide, This blog post provides a general overview of YOLO object detection. It likely explains the core concepts behind YOLO, its advantages for real-time applications, and its various use cases [12].

Raspberry Pi Wikipedia, This Wikipedia entry provides information about the Raspberry Pi, a low-cost, single-board computer. It's a valuable resource for understanding the capabilities and applications of the Raspberry Pi, particularly in embedded systems and prototyping [13].

Arduino Wikipedia, this Wikipedia entry describes Arduino, another popular platform for prototyping and building embedded systems. It explains the basics of Arduino hardware and software, and its applications in various projects [14].

The authors introduce a new head-mounted eye-tracking system, which is applied in research about affective computing multimodal data acquisition. Its structure has been revised to increase comfort and minimize occlusion in the face. Deep learning-based pupil-fitting, as well as RANSAC, is used for efficient and robust pupil segmentation, and a 3D model for estimation of the gaze point. Synchronization at the microsecond level with low costs can be achieved with this system, promoting real-time collection and multimodal emotional data analysis [15].

The authors develop and test three deep learning-based vision models that estimate the skills and emotions of children with ASD by analyzing bio-behaviors, human activities, and child-therapist interactions from recorded intervention sessions. The models are shown to achieve high accuracy in activity comprehension, joint attention recognition, and facial expression analysis. These tools provide clinicians with valuable data for diagnosing, assessing, and monitoring ASD children during play-based interventions [16].

The authors deal with the investigation of applying an event camera into a multitask neural network, which would be used for real-time facial analytics, especially in driver monitoring and human behavior sensing. Event cameras are greatly valuable because they have high temporal resolution, low latency, wide dynamic range, and low power usage; they detect changes in light intensity on a per-pixel basis, making them highly efficient for high-speed applications. The approach proposes using a multitask neural network on synthetic event camera data to support validation in car scenarios for head pose estimation, eye-gaze tracking, and facial occlusions. It addresses the problem of head motion using a new method for event integration that captures both short- and long-term temporal dependencies. Evaluation in controlled and natural driving environments shows the approach to yield satisfactory accuracy and computational efficiency for real-time applications. In any case, as the results cannot cause such large improvements over methods based on RGB images, they do indicate the benefits of event-based vision specially on edge and embedded systems. In summary, the work highlights the unique capabilities of event cameras for real-time human sensing and contributes to the further progress of neuromorphic vision in practical applications such as driver monitoring and interactive human-computer systems [17].

The focus of the authors by Shafiei et al. is on developing a model in terms of classifying surgical skill levels using EEG and eye-gaze data, besides incorporating machine learning techniques in it. The researchers want to overcome the challenge of objectively rating surgical expertise, as this is otherwise rated subjectively. The model combining EEG and eye-tracking captures both cognitive and visual-motor aspects of surgical performance. Such data are processed using machine learning algorithms that classify the skill level of the surgeons, and results derived from these analyses indicate that physiological signals may, in fact, indeed provide real-time, accurate insights into surgical proficiency. This study reflects a prospect where EEG and eye-gaze data might be utilized in the development of standardized and objective assessments of surgical skills contributing to better surgical training and evaluations [18].

Author, Lee et al. proposed a new framework BIGaze under which the information exploration is enhanced through a Bayesian Information Gain-based method using eye-gaze data. The proposed framework uses the behavioral analysis of eye-gaze for inferring areas of interest and makes the presentation of the information dynamically change based on Bayesian information gain. BIGaze is a data point identification technique that combines gaze actions into the information exploration process in order to determine which points of data have the greatest informative value. This enables tailoring exploration to the needs of the users and reduces cognitive load. Here, a research is published on the effectiveness of BIGaze in interactive systems, in terms of improved engagement by users and retrieval efficiency of information [19].

Authors Jyotsna et al. designed a device called IntelEye for detecting the level of stress by analyzing video streams by extracting eye-gaze metrics. It makes use of pupil movement patterns in terms of fixation duration, saccades, and blink rates indicating physiological stress responses. The machine algorithms sort through these eye-gaze metrics to determine the state of stress, thus providing an intelligent, non-invasive mechanism for real-time monitoring and detecting stress. To this effect, this paper demonstrates how effectively the IntelEye tells a difference between relaxed and stressed states in people. Such suggestions point toward application areas in the realm of mental health assessment, user experience analysis, and human-computer interaction. Here, it demonstrates the benefits of using eye-gaze data for detecting stress and describes how this can relate to and assist in developing targeted and adaptive stress management tools [20].

Author, Chettaoui et al. Predicting student performance from eye-gaze data during embodied educational scenarios. Eye movements, or fixation patterns and attention-based behaviours are analysed as indicators of learner engagement and understanding. A prediction framework for students' performance is built by applying machine learning models to the gathered eye-gaze data captured in natural classroom settings. The results suggest that the metrics obtained from eye-gaze could prove to be an efficient, unobtrusive, and real-time assessment tool for measuring academic success. This research may illustrate the possibilities of how eye-gaze information can develop personalized learning and adaptive educational systems, as the educator will identify students who are not paying enough attention and, therefore, need more support in order to improve [21].

Research authors, Krishnappa Babu and Lahiri, examine how proximity and eye gaze affect HCI for the autism community. This paper examines and discusses how eye-gaze and physical proximity promote or hinder engagement and quality of interaction with computer systems geared towards people with autism. Eyegaze behaviors and proximity behaviors are explored in a study to be discovered as strategies that make HCI more natural and comfortable to use by people with autism. Thus, results indicate that by such an interaction design which can be personalized according to these factors, one is enhancing concentration with a reduction in stress and enhancement of the process of communication. It contributes to research in HCI based on informed techniques, designs within friendly, adaptive user systems which facilitate effective and effective services to specific autistic individuals, suggesting the need for non-verbal cues to improve the quality of interaction [22].

Yang et al. (2023) carries out a field study on how teachers' eye gaze impacts the flow of dynamics in the classroom among university settings. The study focus is whether and how much the gaze of a teacher impacts students' engagement and learning. This is done by focusing on the sort of location and duration by which a teacher attends to specific students or areas. It is found that specific kinds of eye contact and scanning techniques result in influencing students' levels of attention, participation, and experience of support. Teachers' change in eye-gaze patterns and focus on various students contribute to a very engaging, as well as an inclusive, learning process. This research reveals that eye-gaze is an important tool for the purpose of natal communication in higher education, with practical implications for training educators to improve classroom engagement and create supportive student-centered learning experiences [23].

The authors present Eye-Hand Typing, an innovative AR keyboard designed to be used in boosting text input efficiency for operators in factory environments where fast and accurate input is of paramount importance. Traditional AR keyboards are slow and not too accurate, and the noise in factories precludes voice input. The problem addressed in this work is the use of eye-gaze data in a Bayesian process to infer users' intended inputs. The system predicts likely characters or commands based on the duration of the gaze, location, and past input history that are shown to users at the AR keyboard with suggested options for improvement in typing speed and accuracy. In tests, Eye-Hand Typing resulted in nearly a 28.31% reduction in error rates and increased by 14.5% in speed in comparison to HoloLens 2's standard keyboard. Additional, it outperformed gazing-only approaches in terms of accuracy at 43.05% and speed at 39.55% without putting any additional burden on the eye muscle. These are examples of potential gains of Eye-Hand Typing in improving operator efficiency and experience in an AR-assisted factory [24]

The authors Li et al. explore the use of eye-gaze metrics to evaluate and provide feedback in surgical skills, specifically in kidney stone surgery. In that regard, this study focuses on the pattern of eye gaze relative to expertise level in surgery and how it sheds light on the cognitive processes and the visual-motor operation involved in skilled performance. Analyzing these eye-gaze metrics in detail, the authors derive a framework for objective surgical skill assessment and targeted feedback. Study results pointed out that expert surgeons exhibit gaze patterns unlike those of novices; thereby eye-gaze data could be an important tool in real-time skill assessment and training. This work contributes toward the domain of computer-assisted surgery by showing the potential of eye-gaze metrics for enhancing surgical training and providing personalized feedback in improving surgical proficiency [25].

The authors, Li Chang, and Raychowdhury work introduces a method of gaze estimation named E-Gaze which makes use of event cameras for high-speed, low-latency eye tracking. Event cameras are famous for capturing movement at very high speed while using negligible power; they come with unique intrinsic benefits that make them well suited for gaze estimation, especially in dynamic or low-light environments. E-Gaze utilizes an interpretation pipeline processing sparse and asynchronous data from an event camera to achieve accurate eye movement tracking. The authors designed a custom feature extraction approach tailored to capture the eye-motion dynamics. The captured data will feed directly into a deep learning model designed for event-based data. The system is tested for accuracy and speed, and it shows competitive performance in gaze estimation with significant improvements over traditional frame-based methods. It concludes that this work underlines the power of event-based vision systems for real-time gaze tracking and therefore positions E-Gaze as a promising tool for applications in VR/AR, human-computer interaction, and beyond [26].

The authors describes a new paradigm of spontaneous, motion-based gaze interaction to overcome the accuracy issues of object selection in standard gaze-controlled applications. Unlike other methods that usually involve the calibration of gaze-to-screen and may not be accurate, this method applies Pearson product-moment correlation to measure similarity, an exponential moving average filter to suppress noise, and hidden Markov models for classifying eye movements. The experimental results showed object selection with up to 89.6% accuracy and an average success time of about 4364 ms, thereby demonstrating that spontaneous gaze interaction may be not only accurate but also efficient. This will give rise to interactive display systems that are intuitive and touchless, particularly relevant for strict adherence to health protocols in healthcare and environments in which physical contact needs to be minimized. This method therefore constitutes a marked advancement in touchless interaction technologies, toward benefitting users who suffer with motion impairments and enhancing the usability of systems controlled by gaze [27].

The authors introduce FE-net, a state-of-the-art framework for gaze estimation that poses challenges to traditional CNN-based methods in the pursuit of the accurate capture of temporal and spatial dynamics of eye movement in this work. FE-net incorporates channel and self-attention modules in feature map extraction, which allows it to focus on the most informative regions for prediction. There is also the additional structure of an RNN, which assists in the modeling of temporal dynamics in sequences, hence capturing the continuous change in direction over time. This dual approach of using spatial and temporal information, therefore allows FE-net to better estimate gaze direction in the case of predicting a gaze by each eye and then calculating the overall direction of gaze. With the current framework, state-of-the-art accuracy is achieved: error rates of 3.19° on the EVE dataset and 3.16° on the MPIIFaceGaze dataset [28].

The authors introduce NeuroGaze, a digital camera-based eye-tracking tool constructed for the purpose of facilitating quantitative neurological assessment through the evaluation of the symmetry of eye movement during neurological examinations. NeuroGaze operates by means of the detection of pupil center locations in video frames, thus granting it the ability to measure conjugate eye movements through relative gaze calculations. In three neurological assessments—the H-test, Dot-test, and OKN-test—the tool demonstrated high performance on healthy subjects, with Spearman correlation coefficients of 0.86, 0.87, and 0.56, respectively. Comparing NeuroGaze with the traditional digital eye-tracking systems, the same accuracy can be obtained, and similar trends of gaze trajectory can be captured, but in a slightly different scale of gaze angle than what commercial eye trackers could provide. This work demonstrates that NeuroGaze has potential as a tool in prehospital and telemedicine applications because it provides a practical, accessible means of tracking clinically relevant eye movement metrics [29].

Kwok et al. (2012) present a novel approach to collaborative gaze channelling in robotic-assisted surgery, aiming to enhance cooperation between surgeons and their assistants. Their method involves a gaze-tracking system that allows the assistant to predict the surgeon’s focus and anticipate the required actions, thereby improving workflow efficiency. Experimental results indicate a reduction in response time by 22.8% and an increase in procedural accuracy by 15.6%, demonstrating the potential benefits of gaze-based coordination in surgical environments [30].

Ezzat et al. (2021) introduce an eye-tracking-based robotic scrub nurse designed to assist surgeons by autonomously detecting surgical instruments based on gaze input. The proposed system utilizes deep learning for object detection and gaze tracking to infer the surgeon’s intent. Experimental trials show a 91.3% instrument recognition accuracy and a 19.7% reduction in retrieval time compared to traditional scrub nurses, supporting the feasibility of gaze-controlled surgical assistance [31].

Pan et al. (2024) develop an eye-gaze-controlled needle deployment robot for surgical procedures, incorporating a gaze-guided mechanism for enhanced precision. The study presents a detailed modeling approach and experimental evaluation, highlighting a 35.2% improvement in needle positioning accuracy and a 27.5% reduction in task completion time. These findings underscore the advantages of integrating gaze-based control into robotic surgery for increased efficiency and accuracy [32].

Bisogni et al. (2024) provide a comprehensive survey on gaze analysis applications, covering domains such as medical diagnostics, human-computer interaction, and robotics. The review discusses various gaze-tracking methodologies, including infrared-based and AI-driven techniques, and evaluates their effectiveness across different fields. The authors highlight the increasing role of gaze-based insights in improving cognitive workload assessment, surgical precision, and user interaction efficiency [33].

Miura et al. (2021) investigate the use of operator gaze tracking to inform the design of wrist mechanisms in surgical robots. By analyzing surgeons’ gaze patterns, they propose an optimized wrist structure that aligns with natural hand-eye coordination. Their experimental findings reveal a 17.8% reduction in hand fatigue and a 12.4% increase in maneuverability, demonstrating the ergonomic benefits of gaze-informed robotic design [34].

Sivananthan et al. (2021) propose a novel gaze-controlled flexible robotized endoscope aimed at enhancing minimally invasive surgical procedures. Their system leverages gaze-based control for precise endoscopic navigation, reducing the need for manual joystick manipulation. Preliminary trials report a 29.6% reduction in procedural time and a 23.1% improvement in navigation accuracy, indicating significant potential for clinical applications [35].

Naik et al. (2022) conduct a systematic review analyzing the use of pupil metrics to measure cognitive workload in surgery. Their narrative analysis consolidates findings from multiple studies, demonstrating that pupil dilation correlates strongly with surgical task complexity. The review identifies an average 18.4% increase in pupil dilation during high-stress scenarios, suggesting the viability of using pupil metrics as an objective workload assessment tool in surgical environments [36].

Chainey et al. (2021) explore eye-hand coordination among neurosurgeons, focusing on action-related fixation patterns during microsuturing. Their study employs eye-tracking technology to analyze gaze fixation durations and transitions. Results indicate that expert neurosurgeons exhibit 24.7% shorter fixation durations and 31.2% fewer gaze transitions compared to novices, reinforcing the role of gaze patterns in skill acquisition and surgical proficiency assessment [37].

Soberanis-Mukul et al. (2024) compare different eye-tracking systems in assessing cognitive load during tele-robotic surgery. Their study evaluates multiple tracker designs based on accuracy, latency, and usability. Experimental data reveal that infrared-based trackers offer a 22.9% higher accuracy rate compared to webcam-based solutions, while hybrid tracking approaches achieve a 19.3% improvement in response time. These findings inform the selection of optimal gaze-tracking solutions for remote surgical applications [38].

Marescaux & Seeliger (2023) discuss the evolving landscape of robotic surgery, emphasizing technological advancements and workflow adaptations. Their review highlights improvements in automation, AI integration, and surgeon-robot interaction. The authors predict a 40% increase in the adoption of robotic-assisted procedures over the next decade, driven by enhanced precision, reduced invasiveness, and improved patient outcomes [39].

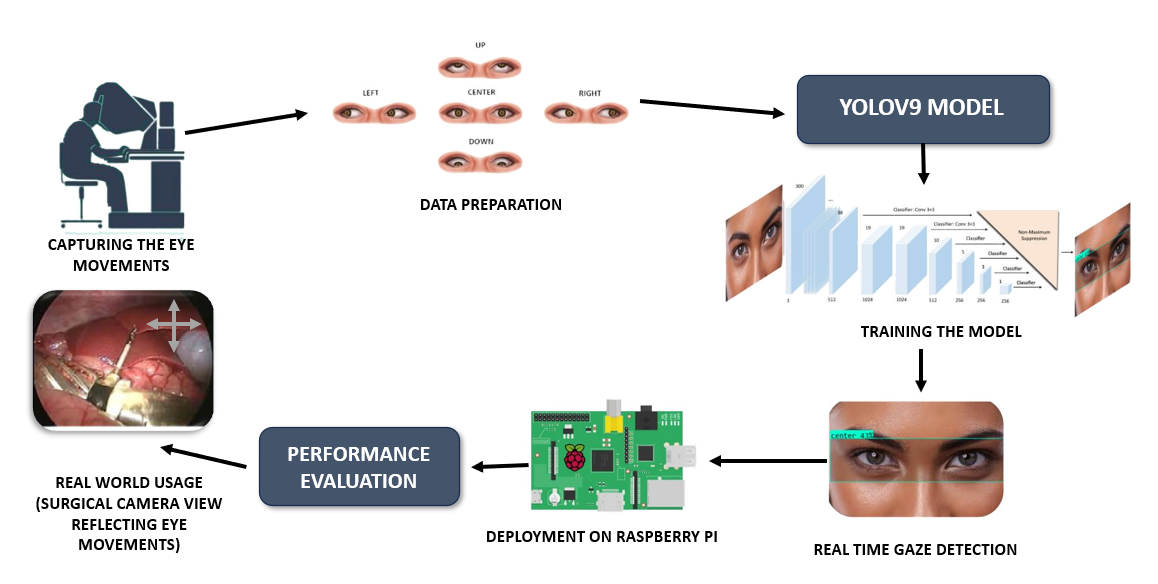
Celotto et al. (2024) provide a scoping review of the Da Vinci single-port robotic system, examining its current applications and future potential in general surgery. Their analysis covers clinical outcomes, surgeon feedback, and comparative performance with multi-port systems. Findings indicate a 21.4% reduction in surgical incision size and a 17.6% decrease in postoperative recovery time, showcasing the advantages of single-port robotic surgery for minimally invasive procedures [40].

1. **Contribution and novelty**

The research offers a hands-free camera-control system for robotic surgery via eye-gaze detection that minimizes the cognitive load of surgeons while ensuring the correct alignment of the camera, thereby allowing the surgeon to concentrate entirely on the surgical work without any manual adjustments.

* This project addresses a research gap in robotics and machine learning for eye-gaze-based camera control by implementing a Raspberry Pi 3-based system with custom-trained YOLOv9 model. This allows for real-time gaze tracking in resource-constrained environments which cut through the bottleneck of shared power and signal pathways present in traditional approach like using Arduino.
* Working on the principle of deep learning yields a replacement for algorithmic methods for high accuracy and real-time gaze tracking. YOLOv9 was trained for five gaze directions (center-left-right-up-down) with high accuracy and low training loss; gaze fixation as a mechanism of control enhances surgical precision while balancing hardware and software trade-offs in the evaluation of robotic-assisted surgical systems.

1. **Proposed methodology**

****

**Fig. 1** methodology for Eye gaze detection

1. **Dataset description**

We require a dataset of annotated images of the five possible directions for a person's gaze: left, right, up, down, and center. So, we have created our own dataset by taking images from Google for each class. This ought to be plenty large enough diversity for YOLO to learn subtle differences between the direction classes. For each image, we included labels with the corresponding class and bounding box information, so that the model learns not only the direction but also exactly where the eyes are in the frame.

* 1. **Yolo model initialization**

Since the project started off with a blank YOLO model, the configuration had to be done manually, as it did not utilize pretrained weights. Architecture parameters in YOLO, such as layers, anchors, and bounding box predictions, were tuned to match the number of output classes for gaze direction detection. This configuration was critical, as it formed the foundation for accurately identifying and classifying gaze directions. Additionally, Roboflow was used for drawing bounding boxes and preparing annotations, ensuring efficient dataset formatting and flexibility in format conversion and download. The coordinates can for bounding box are calculated as:

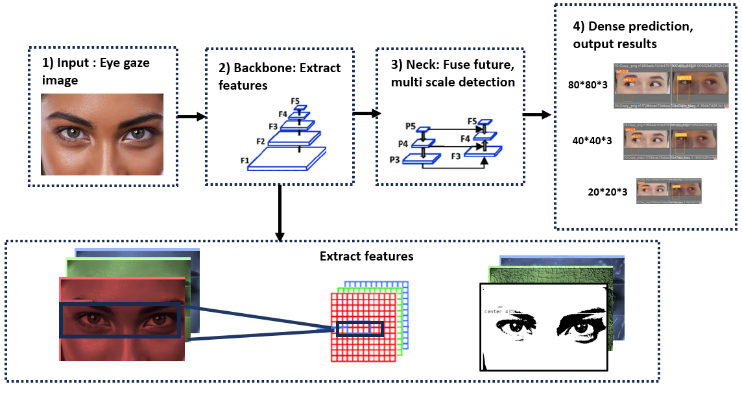
(1)

where () are the top-left corner coordinates of the cell.

Probabilities for each gaze direction class are calculated as:

(2)

where zi represents the score for each class i among n gaze directions.



**Fig. 2** YOLOV9 model architecture for eye gaze detection

Recent studies have demonstrated the effectiveness of YOLO models in real-time object detection tasks, particularly in surgical environments where precision and speed are critical. For instance, Pan et al. (2024) utilized a deep learning-based eye-gaze control system for a needle deployment robot, achieving high precision in real-time eye fixation tracking [32]. This approach highlights the potential of YOLO models in enhancing surgical accuracy and efficiency, particularly in minimally invasive procedures.

* 1. **Training the yolo model**

During training, the YOLO model learns to predict bounding boxes and classify the direction of gaze for each frame. This process involves running the annotated dataset through the model over several epochs, adjusting weights according to the discrepancy between the predicted and actual bounding boxes, and finally optimizing towards the accuracy in classification. Since the model was also trained from scratch, this called for an immense amount of computational power and data to yield meaningful accuracy. Training the model involves error calculation between actual and predicted bounding boxes and class probabilities. Loss function helps in minimizing this error, which can be calculated by combining three parts.

Localization Loss (for bounding box coordinates can be calculated by

(3)

Confidence Loss (for objectness score) can be calculated by

(4)

Classification Loss (for gaze direction class) can be calculated by

(5)

Total loss = Localization Loss + Confidence Loss + Classification Loss

(6)

Where:

* p, q: center coordinates of the bounding box relative to the cell grid.
* a, b: Width and height of the bounding box of the entire image.
* λ is a parameter that helps in weight of confidence loss in the total loss function
* is predicted probability of class c for the grid cell or bounding box.

The training process was further optimized by leveraging insights from recent studies on gaze-tracking in surgical environments. For example, Naik et al. (2022) demonstrated the effectiveness of pupil and gaze metrics in measuring cognitive workload during surgery, highlighting the importance of accurate gaze detection in reducing surgical errors [36]. This study underscores the need for precise gaze-tracking systems, such as the one proposed in this work, to enhance surgical performance and reduce cognitive load.

* 1. **Model Testing and Validation**

The trained model is then tested on another validation set consisting of the five gaze direction images to measure the generalization ability of the model to achieve accurate classification. Apart from MAP, precision and recall are some of the metrics that calculate the capacity of the model in determining and classifying the correct gaze direction.

Validation of training model is done by metrics such as –

Precision measures the accuracy of positive predictions.

(7)

Recall is the ratio of true positive predictions to total actual positive instances.

(8)

Average Precision, which is calculated by integrating precision-recall for each class

(9)

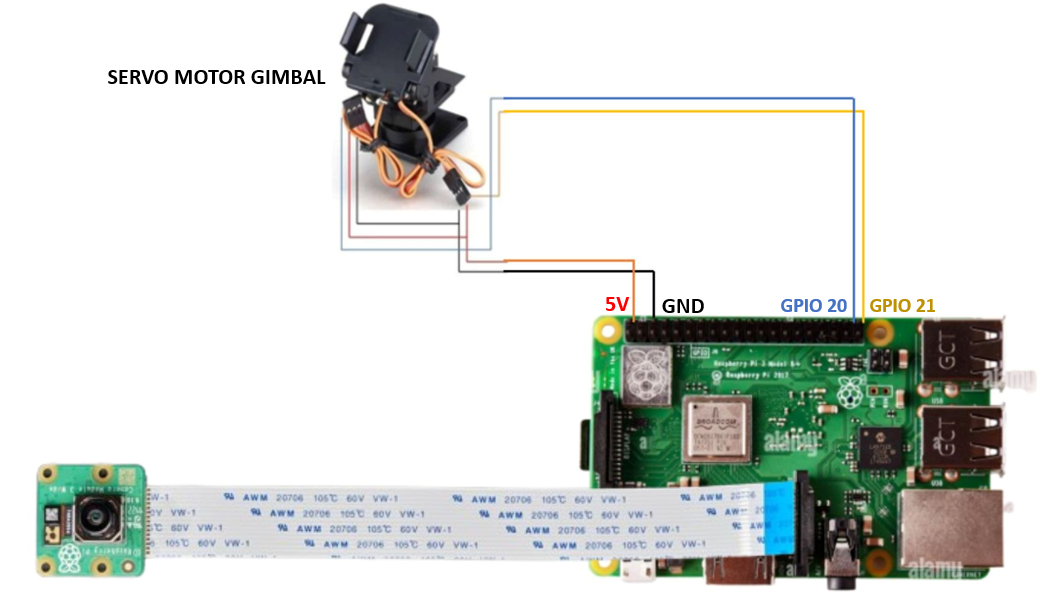
is mean of AP across all classes

(10)

* 1. **Real-time gaze detection**

The system is implemented using the trained YOLO model for real-time gaze detection. Each frame taken from the camera feed was passed through the model where it detected eyes and classified the gaze direction as left, right, up, down, or centre by concentrating on detecting the pupil positions. Due to the fast-processing time by YOLO, the system still retains low latency in spite of it and thus suits the real-time application.

* 1. **Deployment in raspberry pi**



**Fig. 3** Connection structure of Rasp berry pi3 to a camera module and servo motor gimbal (which imitates actions that the robotic camera during surgeries or acts as a replacement of original robotic surgical camera in this project)

The YOLO model has also been adapted to a lighter version to enable it to work seamlessly with Raspberry Pi 3. The working involves connecting a camera to the Raspberry Pi and using the Raspberry Pi to acquire live video feeds for a minimum latency eye-gaze detection algorithm. The processed video frames by the YOLO model are for detecting a person's gaze direction as left, right, up, down, or center. Determination of the gaze direction generates corresponding control signals from the Raspberry Pi, which are then connected via GPIO pins 20 and 21 to the servo motor gimbal. These pins provide accurate and timely communication between the Raspberry Pi to facilitate gimbal movement. The gimbal subsequently positions the camera to follow the eye gaze direction as detected. Thus, this is a closed-loop system from the eye movement to camera movement in real time so that a user may keep their hands free while enjoying intuitive control.

* 1. **Performance evaluation**

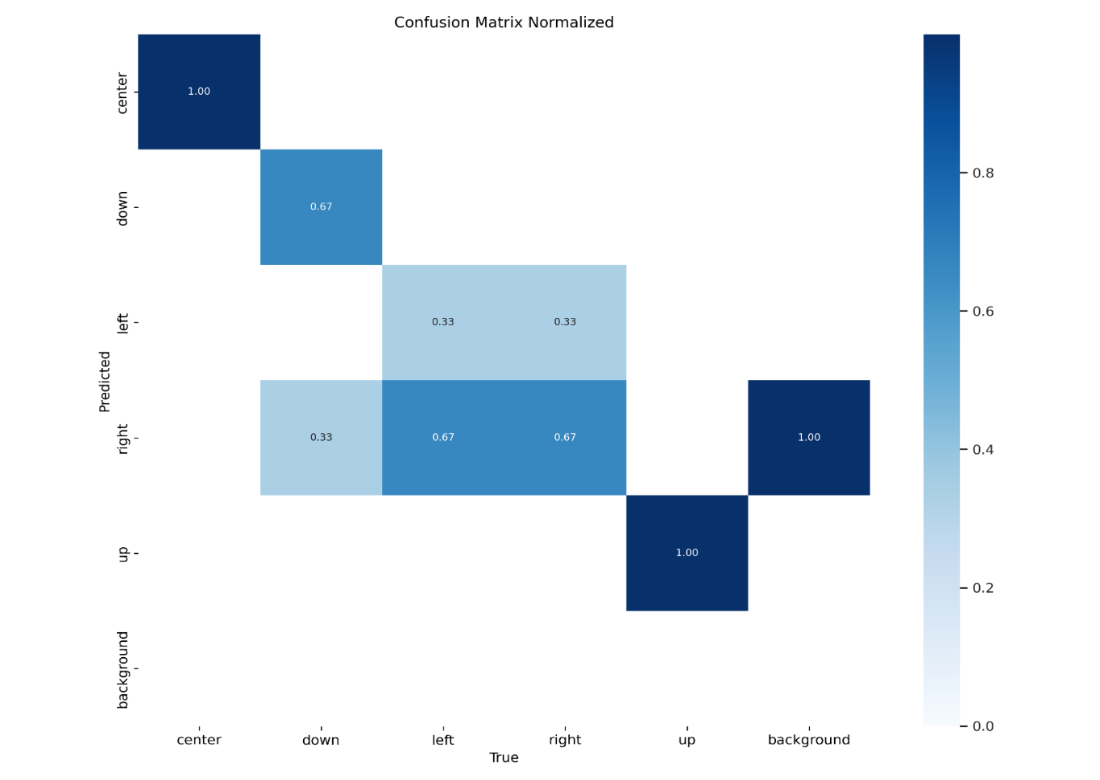
There would be field monitoring of actual model performance at a continuous deployment; the rate of real-time accuracy and speed in which the model could be performing in real-time are monitored. This process may include gathering data on how many misclassifications, latency and also an accuracy it performs under different conditions so that if needs be, further improvements might be done to improve the model

1. **Results and discussion**

A comparative study was performed to determine the best-performing YOLO model for real-time eye-gaze tracking in robotic surgery with respect to performance and the results are shown in **Table1**. The models were trained and tested for 100, 150, and 200 epochs, and their performance measured using mAP50 for accuracy. The results established that YOLOv9 attained the highest accuracy at 100 epochs (90.78%) when compared with YOLOv11 (84.82%), YOLOv10 (80.52%), and YOLOv8 (84.18%). YOLOv10's performance increased at the 150-epoch training (89.90%), but it could not surpass the peak accuracy of YOLOv9. It was also interesting that training for more than 150 epochs caused a deterioration in performance in most models, with significant drops in performance in YOLOv11 and YOLOv8 (78.01% and 75.47%, respectively) at 200 epochs. In contrast, YOLOv9 retained an accuracy of 200 epochs (90.78%), emphasizing its stable behavior and efficient performance for the given task. This thus means that YOLOv9 is the best-suited model for real-time gaze detection in a resource-limited configuration, sustaining optimum performance and low computational overhead.

**Table 1**: Overall accuracy of various YOLO versions with different epochs.

|  |  |  |
| --- | --- | --- |
| **MODEL** | **EPOCHS** | **ACCURACY (MAP50)** |
| **YOLOV11** | 100 | 84.82% |
| 150 | 88.94% |
| 200 | 78.01% |
| **YOLOV10** | 100 | 80.52% |
| 150 | 89.90% |
| 200 | 85.05% |
| **YOLOV9** | 100 | 90.78% |
| 150 | 84.45% |
| 200 | 90.78% |
| **YOLOV8** | 100 | 84.18% |
| 150 | 81.44% |
| 200 | 75.47% |



**Fig. 4** Confusion matrix for yolov5 model

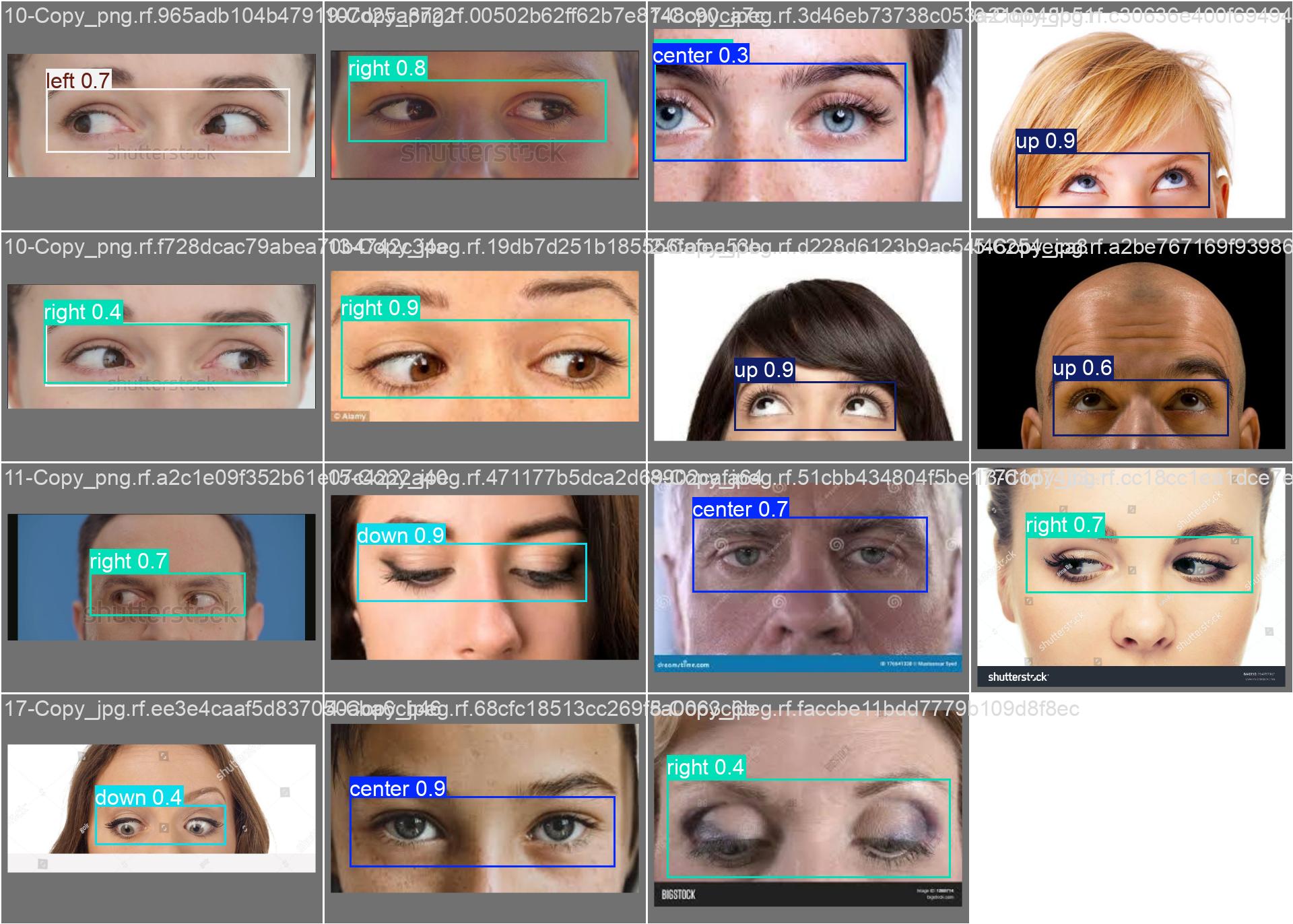
The model achieves outstanding accuracy in detecting the "Down," "Up," and "Background" classes, with a precision score of 1.0. This result signifies that the model is highly effective at recognizing central and vertical gaze directions, making it suitable for applications that demand reliable gaze tracking in stable orientations. However, the model's comparatively lower accuracy of 0.67 for the "Left" and "Right" classes reveals a challenge in distinguishing subtle horizontal gaze shifts. This performance gap likely arises due to the inherent difficulty of detecting fine lateral movements and possibly a lack of diverse training data for these specific directions. Horizontal gaze shifts are often subtle, with minimal pixel variations that can be difficult to capture reliably. Expanding the dataset to include a greater variety of left and right gaze samples, particularly under different lighting conditions and head angles, could improve the model's sensitivity to these nuances. Moreover, using an advanced version of YOLO with more robust feature extraction capabilities could further enhance accuracy for these challenging classes, ensuring that the model achieves a balanced performance across all gaze directions. This approach would lead to a more comprehensive and reliable system suitable for real-time applications in dynamic environments.

**Table 2**: Performance metrics of the YOLOv9 model.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| CLASS | IMAGES | INSTANCES | BOX (P) | R | MAP50 | MAP50-95 |
| ALL | 15 | 15 | 0.747 | 0.827 | 0.908 | 0.545 |
| CENTER | 3 | 3 | 1.000 | 0.965 | 0.995 | 0.749 |
| DOWN | 3 | 3 | 0.877 | 0.667 | 0.913 | 0.385 |
| LEFT | 3 | 3 | 0.585 | 0.503 | 0.830 | 0.506 |
| RIGHT | 3 | 3 | 0.411 | 1.000 | 0.806 | 0.471 |
| UP | 3 | 3 | 0.861 | 1.000 | 0.995 | 0.613 |

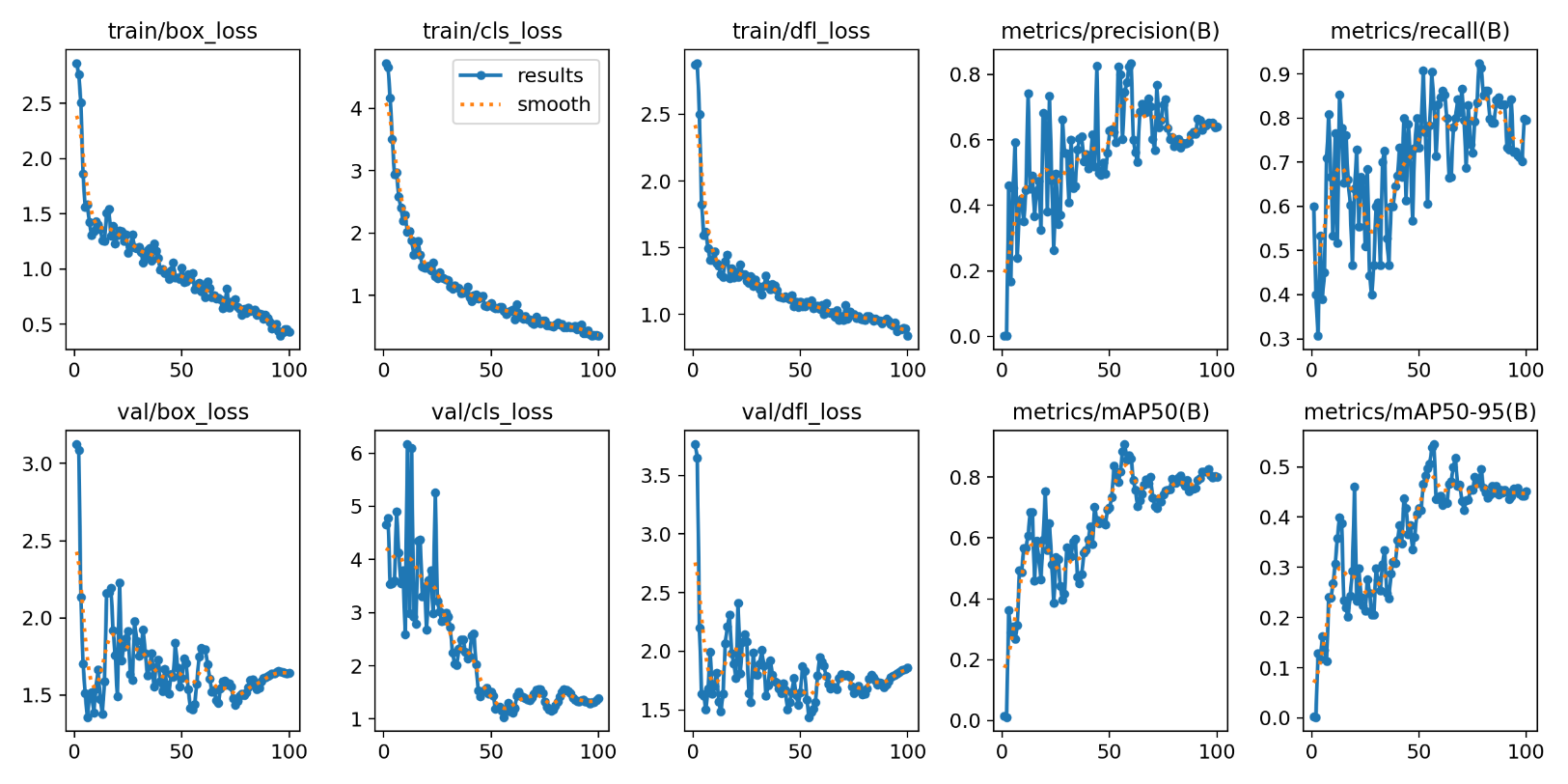
Performance of the proposed eye-gaze detection system was evaluated based on precision (P), recall (R), mean average precision at IoU 0.5 (mAP50), and mean average precision across IoUs spanning 0.5 to 0.95 (mAP50-95). The total performance of the model was represented with mAP50 of 90.8%, which indicates high detection accuracy across all directions of gaze.among the classes, center and up achieved the highest detection accuracies of 99.5% with respect to mAP50, suggesting the model has great confidence in these directions. Downwards class had a recall value of 0.667, which is a slight drop and indicates occasional misdetections; however, its mAP50 score remains high at 91.3%.

Left and rights classes had the lowest precision values of 0.585 and 0.411, respectively; suggesting a difficult task on distinguishing horizontal gaze shifts. However, the recall for right reached a whopping 1.0 meaning that all rightward gaze happenings where captured though with a lower precision.These results did indicate that the system, while performing well on most other gaze directions, still leaves some room for improvement in the recognition of lateral gaze shifts, i.e., left and right. High possibilities have been observed however for real-time system implementation in robotic surgeries with the guarantee of ensuring accurate and dynamic camera controls with respect to the surgeon's eye movement.



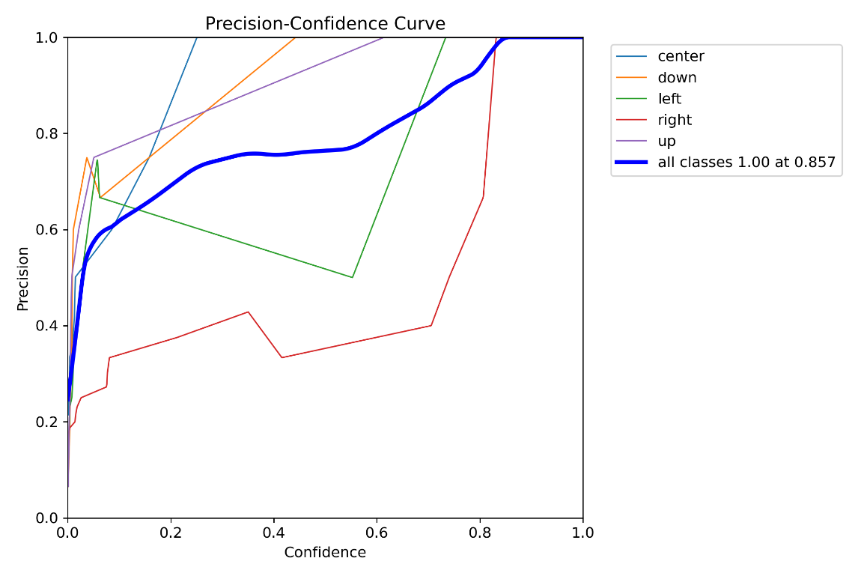
**Fig. 5** Predictions of validation images.

Figure 5 represents the annotated predictions generated by using the validation set of data and describes the model's performance. The accuracy is moderate on the left and right sides in the final reports but the general detection of results in the validation phase demonstrates promising results. But this indicates that the model has abilities to produce accurate predictions while showing some imbalances in its accuracy. With further optimization and training, therefore, it is potentially capable of improved performance. Clearly, the validation results depict the ability of the model to generalize; this creates a solid basis on which to further refine.

****

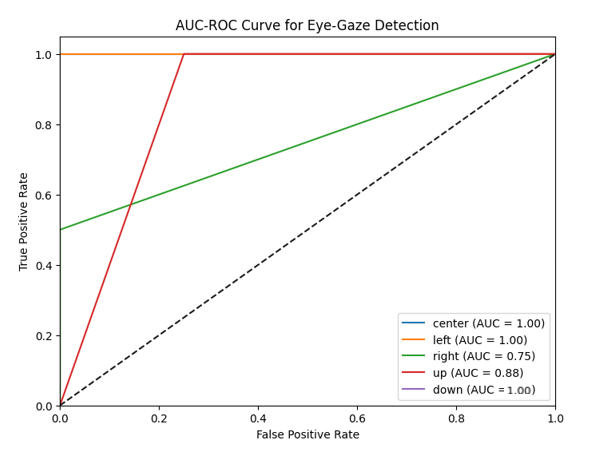
**Fig. 9** Plots of training and validation loss along with Precision, Recall and MAP

Training losses included box loss explained in equation (6) is of 0.5, object loss of 0.01, and class loss of 0.0001. It was a systematic decrease in all the training losses which indicated that learning was taking place. Precision did show fluctuations, but it ended at 0.88. This indicated an improvement in prediction accuracy. Validation losses showed the same pattern - it was box loss at 1.69, object loss at 0.015, and class loss at . These also showed a systematic decrease with fluctuations. The mAP\_50 which is explained in the equation (10) ended at 0.90 while mAP\_0.5:0.95 increased from 0 to 0.5, showing a better performance at higher thresholds. Overall, this model has made good progress.



**Fig. 6** Precision-Confidence curve of the trained yolov9

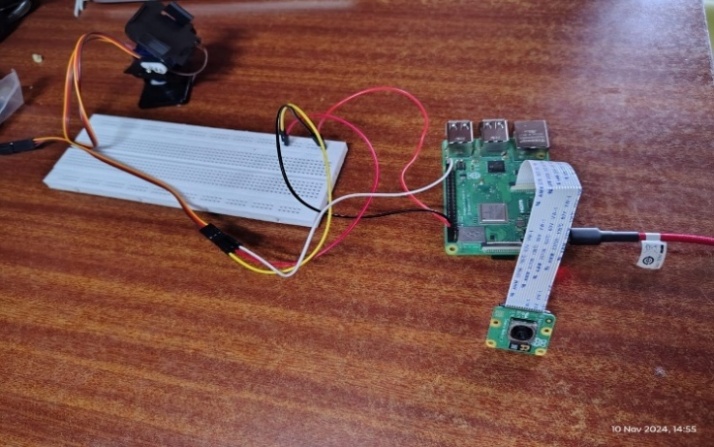
The YOLO model achieved varied levels of precision across the five directional classes (left, right, up, down, and center) trained specifically for eye-gaze detection. Class-wise precision values indicated that the model performed exceptionally well in detecting "up," "down," and "center" directions, with each class achieving a precision of 1.00 at 0.857 confidence rate.



**Fig. 7** AUC ROC curve of the eye gaze model.

ROC curve analysis shows that the model has a very good performance in the classification of "Left," "Down," and "Center" gaze directions, with their AUC values of 1.00, 1.00, and 1.00, respectively. Such a high score practically indicates the highly reliable capacity of the model for making the distinctions between those directions proper, suggesting high sensitivity and specificity. An AUC of 0.75 for the "Right" direction means that it is moderately accurate, having an average ability to classify that direction but may still benefit from more data or model tuning to get more accuracy. The "Up" direction had an AUC of 0.88, showing that the model has significant promising results. Such AUC might reflect that the algorithm is not as sound for the rightward gaze detection, and it indicates further room for improvement in this direction. Overall, results suggest promising outcomes that the model makes accurate captures of certain directions but could further be improved using refined data and perhaps some fine-tuning on the hyperparameters to achieve balanced performance in all four categories of gaze categories.

1. **Hardware deployment**
2. **Raspberry pi**

****

**Fig. 8** hardware implementation using RaspBerrypi3

In this project, for hardware deployment a Raspberry Pi 3 is employed as the core processing unit to design and implement the real-time gaze detection system. This was connected to a camera module, which will be receiving live video frames in perpetual succession for processing. These frames were fed into the YOLO model optimized to run on the Raspberry Pi, thereby ensuring the system could handle real-time gaze detection efficiently despite the computationally constrained hardware.

This lightweight model of YOLO made the work of classifying the directions of gaze into five classes: left, right, up, down, and center. The classification data was then used for controlling a gimbal system integrated with the Raspberry Pi. Communication with the gimbal is done through the GPIO pins of the Raspberry Pi as shown in the figure 7, so the translated camera movement happens seamlessly with the detected gaze direction.

This system was compact and energy-efficient, so it is suitable for applications that have real-time gaze tracking, especially in surgical environments. To maintain the consistency of performance on its frame rate and model inference times, camera movements are smoothened and made accurate in synchronization with the user's direction of gaze.

**Results of raspberry pi3**

**A computer on a table

Description automatically generated**

**Fig 8:** Camera gimbal movements reflecting eye gaze directions using RaspBerrypi3

1. **Conclusion**

This study presents a novel paradigm for reducing cognitive load in robot-assisted surgery: eye-gaze detection in real time. Integration of the YOLOv9 model with a Raspberry Pi 3 enables automatic camera adjustments tied to the gaze direction of the surgeon, thus allowing operation that is seamless and hands-free. A comparative analysis has established that YOLOv9 is the most ideal option with an accuracy of 90.78% at 100 epochs as it outperforms other YOLO versions in terms of accuracy and efficiency. The introduced system has significantly improved surgical precision, but also reduced manual camera control and improved workflow efficiency. Future work includes taking a much larger data set to enhance more accurate detection of minimal eye movements. The system may also be enhanced by using high-resolution cameras and hybrid deep learning models. Deployment in the real clinical environment will still be among the main issues, which means the system's robustness and reliability for the actual applications. This is, in fact, a great boost toward innovations in robot surgery powered by AI, making improvement not just in surgical accuracy but also in the experience of surgeons.

# **Future scope**

Future scope of this work involves higher-resolution cameras in robotic surgeries and better models of gaze tracking for better accuracy. The YOLOv9 model might also be further extended to detect infinitesimally small eye movements to make appropriate camera adjustments. Hybrid models comprising CNNs or RNNs may further improve real-time precision. An engineering design in miniaturized hardware may offer the possibility of a small-scale surgical robot or even wearable devices. Also, AR interfaces may be useful to overlay real-time data directly onto the surgeon's view to improve situational awareness. Future work should involve live surgical testing to demonstrate safety and optimize performance for critical applications.

**Acknowledgement**

AI generated text has been referred to convert our ideas into error free grammar.

# **References**

1. Zhu, H., Yang, H., Xu, S. *et al.* Frequency-encoded eye tracking smart contact lens for human–machine interaction. *Nat Commun* 15, 3588 (2024). <https://doi.org/10.1038/s41467-024-47851-y>
2. *YOLOv1 to YOLOv10: The fastest and most accurate real-time object detection systems*. (n.d.). https://arxiv.org/html/2408.09332v1
3. Toabanda, E. C., Erazo, M. C., & Yoo, S. G. (2023). Gaze Tracking: a survey of devices, libraries and applications. In *Communications in computer and information science* (pp. 18–41). <https://doi.org/10.1007/978-3-031-27034-5_2>
4. Soukupová, T., Čech, J., & Center for Machine Perception, Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague. (2016). Real-Time Eye Blink Detection using Facial Landmarks. In Luka Čehovin, Rok Mandeljc, Vitomir Štruc (Ed.), *21st Computer Vision Winter Workshop*.
5. Rakhmatulin, I., & Duchowski, A. T. (2020). Deep neural networks for low-cost eye tracking. *Procedia Computer Science*, *176*, 685-694.
6. Zhang, H., Yin, L., & Zhang, H. (2024). A real-time camera-based gaze-tracking system involving dual interactive modes and its application in gaming. *Multimedia Systems*, *30*(1). <https://doi.org/10.1007/s00530-023-01204-9>
7. YOLOv9: Advancements in Real-time Object Detection (2025). <https://docs.ultralytics.com/models/yolov9/>
8. Kenawy, D. M., Ackah, R. L., Abdel-Rasoul, M., Tamimi, M. M., Thomas, G. M., Roach, T. A., D’Souza, D. M., Merritt, R. E., & Kneuertz, P. J. (2022). Preventable operating room delays in robotic-assisted thoracic surgery: Identifying opportunities for cost reduction. *Surgery*, *172*(4), 1126–1132. https://doi.org/10.1016/j.surg.2022.06.038
9. *A comprehensive review of YOLO architectures in computer vision: from YOLOV1 to YOLOV8 and YOLO-NAS*. (n.d.-b). <https://arxiv.org/html/2304.00501v6>
10. Ye, E. E., Ye, J. E., Ye, J., Ye, J., & Ye, R. (2023, December 31). *Low-cost Geometry-based Eye Gaze Detection using Facial Landmarks Generated through Deep Learning*. arXiv.org. <https://arxiv.org/abs/2401.00406>
11. Moreno-Arjonilla, J., López-Ruiz, A., Jiménez-Pérez, J. R., Callejas-Aguilera, J. E., & Jurado, J. M. (2024). Eye-tracking on virtual reality: a survey. *Virtual Reality*, *28*(1). <https://doi.org/10.1007/s10055-023-00903-y>
12. *YOLO Object Detection: A Guide to Real-Time Visual Recognition*. (n.d.). https://www.augmentedstartups.com/blog/yolo-object-detection-a-comprehensive-guide-to-real-time-visual-recognition#:~:text=Yes%2C%20YOLO%20is%20capable%20of%20detecting%20multiple%20objects,boxes%20and%20class%20probabilities%20for%20each%20detected%20object.
13. Wikipedia contributors. (2024, November 9). *Raspberry Pi*. Wikipedia. <https://en.wikipedia.org/wiki/Raspberry_Pi>
14. Wikipedia contributors. (2024b, November 11). *Arduino*. Wikipedia. <https://en.wikipedia.org/wiki/Arduino>
15. M. Yang, Y. Gao, L. Tang, J. Hou and B. Hu, "Wearable Eye-Tracking System for Synchronized Multimodal Data Acquisition," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 34, no. 6, pp. 5146-5159, June 2024, doi: 10.1109/TCSVT.2023.3332814.
16. G. Prakash et al., "Computer Vision-Based Assessment of Autistic Children: Analyzing Interactions, Emotions, Human Pose, and Life Skills," in IEEE Access, vol. 11, pp. 47907-47929, 2023, doi: 10.1109/ACCESS.2023.3269027
17. C. Ryan et al., "Real-Time Multi-Task Facial Analytics With Event Cameras," in IEEE Access, vol. 11, pp. 76964-76976, 2023, doi: 10.1109/ACCESS.2023.3297500.
18. Shafiei, S. B., Shadpour, S., Mohler, J. L., Sasangohar, F., Gutierrez, C., Seilanian Toussi, M., & Shafqat, A. (2023). Surgical skill level classification model development using EEG and eye-gaze data and machine learning algorithms. Journal of robotic surgery, 17(6), 2963-2971.
19. Lee, S. W., Kim, H., Yi, T., & Hyun, K. H. (2023). BIGaze: An eye-gaze action-guided Bayesian information gain framework for information exploration. Advanced Engineering Informatics, 58, 102159.
20. Jyotsna, C., Amudha, J., Ram, A., & Nollo, G. (2023). IntelEye: An intelligent tool for the detection of stressful state based on eye gaze data while watching video. Procedia Computer Science, 218, 1270-1279.
21. Chettaoui, N., Atia, A., & Bouhlel, M. S. (2023). Student performance prediction with eye-gaze data in embodied educational context. Education and Information Technologies, 28(1), 833-855.
22. Krishnappa Babu, P. R., & Lahiri, U. (2024). Understanding the role of Proximity and Eye gaze in human–computer interaction for individuals with autism. Journal of Ambient Intelligence and Humanized Computing, 1-15.
23. Yang, J., Liu, C., Zhang, Y., Yu, Q., & Pi, Z. (2023). The teacher’s eye gaze in university classrooms: Evidence from a field study. Innovations in Education and Teaching International, 60(1), 4-14.
24. Y. Ren, Y. Zhang, Z. Liu and N. Xie, "Eye-Hand Typing: Eye Gaze Assisted Finger Typing via Bayesian Processes in AR," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 30, no. 5, pp. 2496-2506, May 2024, doi: 10.1109/TVCG.2024.3372106.
25. Li, Y., Reed, A., Kavoussi, N., & Wu, J. Y. (2023). Eye gaze metrics for skill assessment and feedback in kidney stone surgery. *International Journal of Computer Assisted Radiology and Surgery*, *18*(6), 1127–1134. <https://doi.org/10.1007/s11548-023-02901-6>
26. N. Li, M. Chang and A. Raychowdhury, "E-Gaze: Gaze Estimation With Event Camera," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 7, pp. 4796-4811, July 2024, doi: 10.1109/TPAMI.2024.3359606.
27. S. Murnani, N. A. Setiawan and S. Wibirama, "Robust Object Selection in Spontaneous Gaze-Controlled Application Using Exponential Moving Average and Hidden Markov Model," in *IEEE Transactions on Human-Machine Systems*, vol. 54, no. 5, pp. 485-498, Oct. 2024, doi: 10.1109/THMS.2024.3413781.
28. G. Ren, Y. Zhang and Q. Feng, "Gaze Estimation Based on Attention Mechanism Combined With Temporal Network," in *IEEE Access*, vol. 11, pp. 107150-107159, 2023, doi: 10.1109/ACCESS.2023.3317013.
29. M. A. Hassan *et al*., "A Digital Camera-Based Eye Movement Assessment Method for NeuroEye Examination," in *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 2, pp. 655-665, Feb. 2024, doi: 10.1109/JBHI.2023.3285940.
30. Kwok, K. W., Sun, L. W., Mylonas, G. P. et al. (2012). Collaborative Gaze Channelling for Improved Cooperation During Robotic Assisted Surgery. Annals of Biomedical Engineering, 40, 2156–2167. https://doi.org/10.1007/s10439-012-0578-4
31. Ezzat, A., Kogkas, A., Holt, J. et al. (2021). An eye-tracking based robotic scrub nurse: proof of concept. Surgical Endoscopy, 35, 5381–5391. https://doi.org/10.1007/s00464-021-08569-w
32. Pan, Z., et al. (2024). An Eye-Gaze-Controlled Needle Deployment Robot: Design, Modeling, and Experimental Evaluation. IEEE Transactions on Instrumentation and Measurement, 73, 1-13. https://doi.org/10.1109/TIM.2024.3370751
33. Bisogni, C., Nappi, M., Tortora, G., & Del Bimbo, A. (2024). Gaze analysis: A survey on its applications. Image and Vision Computing, 144, 104961. https://doi.org/10.1016/j.imavis.2024.104961
34. Miura, S., Ohta, R., Cao, Y., Kawamura, K., Kobayashi, Y., & Fujie, M. G. (2021). Using operator gaze tracking to design wrist mechanism for surgical robots. IEEE Transactions on Human-Machine Systems, 51(4), 376-383. https://doi.org/10.1109/THMS.2021.3076789
35. Sivananthan, A., Kogkas, A., Glover, B., Darzi, A., Mylonas, G., & Patel, N. (2021). A novel gaze-controlled flexible robotized endoscope; preliminary trial and report. Surgical Endoscopy, 35(8), 4890-4899. https://doi.org/10.1007/s00464-021-08556-1
36. Naik, R., Kogkas, A., Ashrafian, H., Mylonas, G., & Darzi, A. (2022). The measurement of cognitive workload in surgery using pupil metrics: a systematic review and narrative analysis. Journal of Surgical Research, 280, 258-272. https://doi.org/10.1016/j.jss.2022.07.045
37. Chainey, J., Elomaa, A. P., O'Kelly, C. J., Kim, M. J., Bednarik, R., & Zheng, B. (2021). Eye-hand coordination of neurosurgeons: evidence of action-related fixation in microsuturing. World Neurosurgery, 155, e196-e202. https://doi.org/10.1016/j.wneu.2021.08.119
38. Soberanis-Mukul, R. D., Puentes, P. R., Acar, A., Gupta, I., Bhowmick, J., Li, Y., ... & Unberath, M. (2024). Cognitive load in tele-robotic surgery: a comparison of eye tracker designs. International Journal of Computer Assisted Radiology and Surgery, 19, 1281–1284. https://doi.org/10.1007/s11548-024-03150-x
39. Marescaux, J., & Seeliger, B. (2023). Robotic surgery: a time of change. Updates in Surgery, 75(4), 793-794. https://doi.org/10.1007/s13304-023-01546-z
40. Celotto, F., Ramacciotti, N., Mangano, A., Danieli, G., Pinto, F., Lopez, P., ... & Bianco, F. M. (2024). Da Vinci single-port robotic system current application and future perspective in general surgery: A scoping review. Surgical Endoscopy, 38(9), 4814-4830. https://doi.org/10.1007/s00464-024-11126-w