International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *24 August 2023* ***Revised:*** *09 October 2023* ***Accepted:*** *18 October 2023*

Pupil Position by an Improved Technique of YOLO Network for Eye Tracking Application

S. Mary Rexcy Asha[1], Dr. G. Victo Sudha George[2], Dr. V. Gokula Krishnan[3]
[1]Research Scholar, Department of CSE, Dr. M.G.R. Educational and Research Institute, Maduravoyal
Chennai, Tamil Nadu 600 095, India
Email: rexcyasha@gmail.com
[2]Professor, Department of CSE, Dr. M.G.R. Educational and Research Institute, Maduravoyal
Chennai, Tamil Nadu 600 095, India
Email: victosudhageorge@drmgrdu.ac.in
[3]Professor, Department of CSE, Saveetha School of Engineering,
Saveetha Institute of Medical and Technical Sciences, Thandalam, Chennai, Tamil Nadu, India
Email; gokul_kris143@yahoo.com

Abstract— This Eye gaze following is the real-time collection of information about a person's eye movements and the direction of their look. Eye gaze trackers are devices that measure the locations of the pupils to detect and track changes in the direction of the user's gaze. There are numerous applications for analyzing eye movements, from psychological studies to human-computer interaction-based systems and interactive robotics controls. Real-time eye gaze monitoring requires an accurate and reliable iris center localization technique. Deep learning technology is used to construct a pupil tracking approach for wearable eye trackers in this study. This pupil tracking method uses deep-learning You Only Look Once (YOLO) model to accurately estimate and anticipate the pupil's central location under conditions of bright, natural light (visible to the naked eye). Testing pupil tracking performance with the upgraded YOLOv7 results in an accuracy rate of 98.50% and a precision rate close to 96.34% using PyTorch.

Keywords- Eye Gaze Tracking, Pupil Position, Iris Center Localization, Visible Light Mode, YOLO Network.

I. INTRODUCTION

Our ability to perceive the world around us is primarily based on our ability to see and hear. Humans may move their eyes to focus on the most important features in the environment, whereas ears are fixed in place. Eye trackers, the instruments used to perform this measurement, are usually referred to as "eye trackers." It is possible to obtain information about the subject's eye movement and its relationship to the subject's head using an eye tracker [1-2]. Monitoring vision, one of the most critical senses for controlling grasping and movement, is now possible with wearable eye gaze trackers [5-6]. Hence, the use of eye gaze followers can be beneficial in the restoration and enhancement of functionality in both the upper and lower limbs. Humans use visual input to direct and guide hand drive through a mechanism known as hand–eye coordination. A person's gaze provides guidance on how to approach and grab an object during grasping tasks [7-8]. The object a person is attempting to grip can be detected using eye gaze tracking and object recognition. A patient's intention during rehabilitation can also be detected and supported using gaze information [9]. The use of head-mounted eye gaze trackers and scene cameras in hand prostheses and orthoses can therefore improve these devices [10-12]. Visual input is a primary sensory input utilized to govern locomotion, since it plays a critical role in trajectory preparation, stabilizes the body, and is vital for fall anticipation [13-15].

When used in conjunction with other assistive technologies, such as eye gaze monitoring, it has the potential to make a significant difference in the lives of people with disabilities [16, 17]. Eye gaze following info can be used to control assistive devices like wheelchairs more precisely [18-19]. People with limited functional regulator of the legs (such as those with multiple sclerosis, Parkinson's disease can gain some independence using eye gaze tracking in conjunction with other sensor data, such as inertial measurement units (IMUs). A patient's eye movements can now be tracked as they interact with a computer, browse the Internet, and read e-books [20].

Eye gaze following refers to the process of determining where a person is focusing their attention. [21] The term "point of vision" describes this precise location in space. It's been used, for example, to examine how people interact with computers and scan patterns psychologically. There are two kinds of eye gaze tracking systems mounted eye trackers can be moved around, whereas remote eye trackers remain in one place. Metal contact lenses, as employed in early eye trackers, have been replaced by infrared cameras and bright or dark pupil approaches [22] in modern eye trackers. In order to find the center of the pupil, these techniques are employed. The tracker uses the corneal reflection and the pupil to determine the location of the target on the screen at which the individual

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 24 August 2023 **Revised:** 09 October 2023 **Accepted:** 18 October 2023*

is focusing their attention. More expensive than infrared eye trackers, high-speed video camera-based eye trackers provide more accurate results than webcam-based eye trackers. Such eye trackers use deep learning and computer vision technologies to measure eye movement.

A network-based deep-learning architecture, YOLOv7, is used in this study to improve pupil tracking estimation, making it more accurate. The deep-learning network model provided as the pupil detector is well-trained. There are less light reflection and shadow interference errors in visible light mode with the suggested method than with the prior designs, and this improves accuracy of eye tracking system calibration as a whole. The wearable eye tracker's gaze tracking capability will therefore be enhanced in low-light circumstances.

Those works that are relevant to the research project are used in Section 2. Section 3 provides a brief explanation of the suggested model. Section 4 discusses how the proposed model may be tested using existing methods. In Section 5, the conclusion is given.

II. LITERATURE REVIEW

Computer vision and pattern recognition are two areas where deep-learning technology has made a significant impact in recent years. Real-time object detection and recognition can be achieved with deep learning technology.

When using appearance-based gaze estimate approaches, just using a single camera will limit the application area to a short distance. Fortunately, the authors [23] devised a novel long-distance gaze measurement method to solve this problem. The LSC eye tracker (LSC eye tracker) employed a commercial eye tracker to obtain gaze data, and a long-distance camera to capture face appearance images. Deep CNN models were used to learn between appearance images and gazes during training. Based on entrance photos captured by a single camera and trained CNN simulations, the LSC eye tracker was able to accurately predict gazes in the application phase.

In [24], the authors used a deep learning-based strategy to offer a successful eye-tracking solution. YOLOv3 was used to regulate the user's gaze position while OpenCV was used to regulate the user's face's location in relation to the computer's infrared LED. An inexpensive and accurate remote eye-tracking device was built in [25] by means of a smartphone industrial prototype equipped with a camera and infrared lights. While the head and the device were free to move, we used a 3D gaze-estimation model to precisely approximation the point of gaze (PoG). To precisely determine the input ocular properties, the method used CNN. For smartphones, the hybrid method featured artificial lighting and a three-dimensional gaze estimate model in addition to the CNN feature extractor that is now used.

It was shown in [26] that a CNN can reliably segment entire elliptical structures, even when occluding objects, while also providing better tracking of the pupil and iris centers within a two-pixel error margin than standard segmentation of eye parts for multiple publicly available synthetized eyes.

Human-Computer Interaction (HCI) gets a nonverbal communication boost in [27], thanks to an eye-view-based nonverbal communication paradigm. Traditional gaze detection systems offer remarkable performance and endurance. In any case, there is a need to update these systems. As a result, they can only be used in the lab and are difficult to put into practice in the real world. We recommend utilizing a webcam to track the person's gaze in this situation. An effective visual monitoring framework based on models is provided using a webcam's 2D coordinates. The platform's work is to make HCI easier and, as a result, improve the usefulness of technology and the privacy of users. Implicit human gaze patterns on displayed items were successfully used to contact people's intentions in this experiment. Furthermore, research has demonstrated that it is easy to understand and employ eye contact. In addition, a specific reference system should be used to calculate the subject-dependent ocular parameters. Finally, a monitoring and interpretation system for implicit communication has been created. The technology is able to identify the activities and requirements of the home environment after recognizing the user's implicit purpose to support through the act of the eye-gaze. Finally, the implied goal can be employed to inform caregivers of the proper service to provide.

In [28], an auto-calibration approach for 3D gaze estimation is proposed. An RGBD camera is used as the scene camera of our system in order to gather the precise 3D structure of the surroundings. Saliency maps can be obtained from scene photos using the saliency detection method, and 3D salient pixels in the scene are taken into account as potential 3D calibration targets. It is developed using eye pictures to determine gaze vectors for the 3D model. Our calibration method achieves auto-calibration by merging 3D conspicuous pixels and gaze vectors. As a final step, gaze vectors are calibrated and the point cloud is formed using the RGBD camera. Experiments have shown that the suggested system can achieve indoors an average precision of 3.7 centimeters and outdoors an average precision of 4 centimeters. It is also possible to track users' in real scenes using the proposed system's improved depth measurement.

An operating room needle position and orientation might be precisely realized using a tiny robotic guiding system in [29]. The precision of needle placement during interventional therapy is being evaluated using experimental investigations based on the Robot Operating System (ROS). We can achieve a robot's end effector distance error to the target point within 1mm using our proposed robotic hardware and an eye gaze-based control system.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 24 August 2023 **Revised:** 09 October 2023 **Accepted:** 18 October 2023*

It was proposed in [30] that a gaze detection approach based on an algorithm named Auto-Keras might be used to autonomously create the neural architecture. The Columbia Gaze Data Set is used to build a neural network. In order to validate our model's generalizability, we run it through a series of tests in an online environment. Instead of relying on facial landmarks and filters, the suggested method instead makes use of geometrical operators such as morphological operators and dib facial landmarks to better capture the subject's gaze.

In [31], the authors proposed a method that uses isophote properties to find the center of the iris and cylindrical parameters to estimate where the eyes are. To get accurate gaze directions, these estimated locations are used to model and compare with reference positions. The author came up with a way to estimate eye gaze that takes head pose and eye location into account. The integrated model is better than separate models that were used to estimate head pose and eye location, and the average error is between 2 and 5 degrees.

In [32], the authors showed how high-dimensional 3D cameras and wearable sensors can be used to track a person's gaze and detect their face in real time. The system can be made even better by combining gaze tracking with techniques for figuring out how people feel. This will improve the game interface by giving feedback to the system that was designed.

In [33], the authors describe an EGT system that expands the visible range of the eye-gaze estimator. In particular, the method uses a good model to estimate the corners of the user's eyes and combines the learned data to estimate the iris. The proposed system can deal with big changes in where the eye is looking and do calculations quickly. This proposed system can do calculations quickly and can handle big changes in eye position. The authors came up with the idea of using a wearable device to track how the head moves while playing sports. For field applications, like sports, the wearable device needs to be able to track the user's gaze and have lighting that doesn't change when the user moves.

III. PROPOSED SYSTEM

The proposed wearable eye tracking method is divided into several steps as follows:

(1) Collecting the visible-light near field eye images in datasets and labelling these eye image datasets;

(2) Choosing and designing the deep-learning network architecture for pupil object detection;

(3) Training and testing the trained inference model;

(4) Use deep learning model to infer and detect the pupil’s object, and estimate the centre coordinate of pupil box;

3.1. Dataset description

Each of the 25 subjects in the ARGaze dataset generated 1,321,968 eye gaze photographs. Now you can access it for free at https://doi.org/10.17605/OSF.IO/CJXYN15, the repository where it was originally stored. The photos of the dataset are arranged in hierarchical directories for each participant and each experimental scene (samples are provided in Figure 1). Images of the participants' eyes and their related histograms can be found under the subdirectory histogram (Fig. 3). A 1280 x 720 pixel image sequence from the globe camera is used to create video clips in the scene preview subdirectory. As a result, these video snippets are shorter than their job durations due of the data cleaning process. Most sequences containing 26,552 photos have a video clip length of 7 minutes, 23 seconds (roughly 60 frames per second, which was the original frame rate throughout the session). In the metadata.xlsx file, the theoretical length of preview videos is stored. Eye-tracking photographs are organized by participant number in a series of image directories (e.g., P1). P1 S1 directories are created for each experimental scene within each image directory.

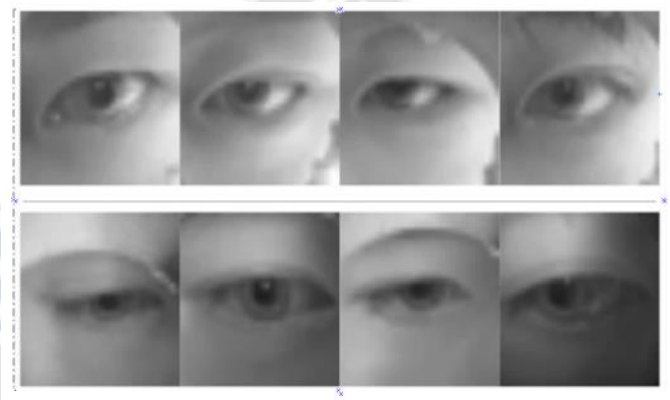

Figure 1: Sample dataset vide* converted image.

P1 has two subdirectories, P1 S1 and P1 S2, for instance. Scenes 1 (S1) and 2 (S2) relate to augmented and real-world scenes. PNG files with a resolution of 32x32 pixels and an 8-bit color depth are used to create the eye gazing images, which are labelled with timestamps to make it easy to match them up with the associated eye images and real-world views. Eye gaze images in a high-resolution encoded.mp4 video file can also be found in the same location.

3.2. Basic Characteristics of YOLOv7 Network

Fast, accurate, and lightweight are all hallmarks of the YOLOv7 network, which is the company's newest offering. YOLOv7 has four primary models: the extended model YOLOv7x, the benchmark model YOLOv7l, and the preset simplified versions YOLOv7s and YOLOv7m. In general, model parameters, as well as the number of convolution kernels deployed across the network, are both lowered.

Table 1 show the YOLOv7 network topology, which comprises of the Input, Backbone, Neck, and Head networks. As a result, the Input terminal uses Mosaic data augmentation, adaptive anchoring, adaptive image scaling, and other

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *24 August 2023* ***Revised:*** *09 October 2023* ***Accepted:*** *18 October 2023*

advanced features. Images are aggregated and features are formed using a CNN known as the Backbone network. In addition to the focus module, CBL module, and other modules, it consists mostly of the CBL module. In order to construct FPN and PAN, the Neck network combines layers of feature aggregation for mixed and combined picture information. In addition to the CBL, Upsample, and CSP2 X modules, it also includes other modules. When computing the bounding box loss, the Head terminal uses GoU_ Loss.

Table 1: YOLOv7 network structure

YOLOv7	**Features**
Neck	CBL, 5×CSP2_X, Upsample, Concat, PAN+FPN
Input	Mosaic data augmentation, adaptive image scaling
Head	GIoU_Loss
Backbone	Focus, CBL, 3×CSP1_X, SPP

3.3 Improvement of Proposed YOLOv7 Network

Deep learning's industrialization is hampered by the technology's enormous computational demands. Optimizing object identification methods for grasping robots is a high concern for minimizing the amount of computation and storage space required. Because of this, our model optimization technique is to use the network pruning method to reduce the size of the YOLOv7 model and make it more efficient. By lowering pruning method improves generalization performance and prevents over-fitting. Figure 2 depicts our revised YOLOv7 network design, which we provide in this research.

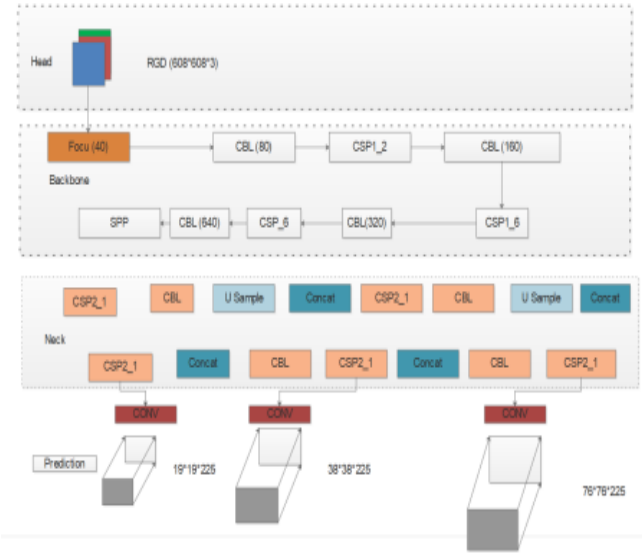

Figure 2: Architecture of improved YOLOv7 network

There are four primary pieces to the enhanced network, which may be seen in Figure 1. The Input terminal receives the collected data, Neck networks are used to prune the network, and Forecast terminals are used to deliver model predictions.

The Focus module is used to speed up the training process. It has the following structure: To begin, the Input terminal used the slice operation to divide the three-channel image dataset into four slices, each measuring 304 304 3. Second, a feature map was generated by concatenating the four depth slices (the image size was 304 304 12). It was then utilized to create a new feature map using a convolution layer that contained 40 convolution kernels (the image's dimensions were 308x304). Final results were created using activation function, and then sent to CBL unit.

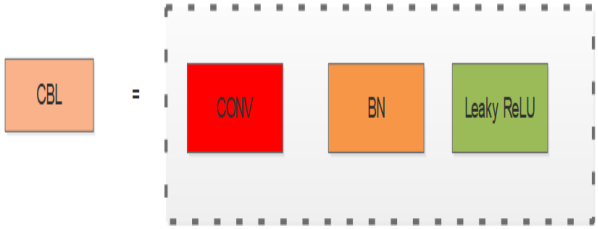

Figure 3: Structure of the CONV-BN-Leaky ReLU (CBL) module

Convolution kernels, BN layers, and leaky ReLU activation functions are all components of CBL module (see Figure 3), which is a small component in YOLO's backbone and neck networks. The number of convolution kernels determines the size of CBL module's output image, which is determined by how many BN layers and ReLU activation functions are used.

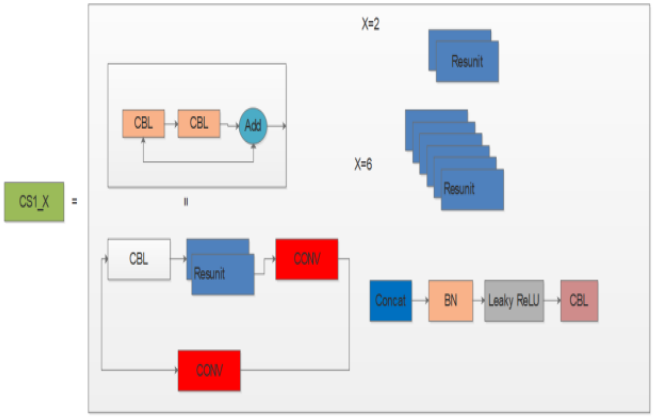

Figure 4: Structure of the CSP1_X module.

This is the CSP1 X module, which is part of the backbone network's third level (seen in Figure 4). Based on the CSPNet concept, the CSP1 X and CSP2 X modules were developed using the same principles. To save time and ensure accuracy, the module divides the basic layer's feature mapping into two sections before combining them using a cross-stage hierarchical structure. Make up the CSP1 X module, which tries to better extract the image's deep information. Its output is the CBL modules and the original input, which is

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *24 August 2023* ***Revised:*** *09 October 2023* ***Accepted:*** *18 October 2023*

represented by the number of Resunits that are present. The CSP1 X module's structure is as follows: First, the initial input was divided into two branches, and the appropriate convolution operation was carried out in each branch. Concatenation of the output feature maps from the two branches was used to connect them. After that, we did batch normalization (BN) and processing on the leaky ReLU activation function. Finally, the CBL component completed the convolution, resulting in an output feature map with the same size as the CSP1 X module's original input feature map.

When it comes to backbone networks, the Spatial Pyramid Pooling (SPP) module is the ninth layer, and its goal is to enhance the network's ability to receptively receive information from arbitrary resolution feather maps. The following is its specific structure: First and foremost, the CBL module performed the convolution operation. Second, three parallel maximum pool layers were used to conduct the maximum pooling procedure. Finally, the feature map after maximum pooling and the feature map after convolution had a strong connection. Finally, the CBL module was used to conduct the convolution procedure once more.

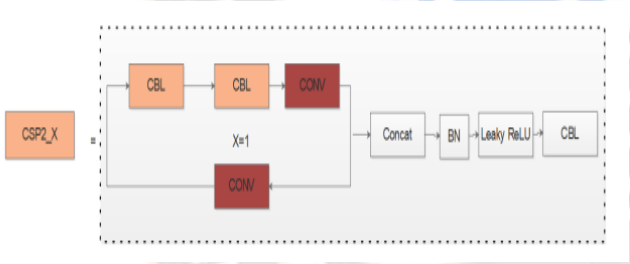

Figure 5: Structure of CSP2_X module.

The CSP2 X module is located at the very bottom of the Neck network (seen in Figure 5). The CSP2 X module has a similar structure to the CSP1 X unit, with the sole variation being that X in the CSP2 X unit signifies the sum of CBL components. Neck network CSP2 X modules are all CSP2 1 in the upgraded YOLOv7 architecture.

3.3.1. Improved YOLOv7 Network Strategy

A hidden layer pruning techniques used for fine-tuning the depth of the network, and a convolution kernel pruning method is used for fine-tuning its breadth in this paper. For network depth, we apply a hidden layer pruning strategy that reduces CSP structure's number of residual components; the network depth for the revised YOLOV7 model, the second has two, and the third structure has six. Five CSP2 modules are used in one residual component. It's possible to reduce YOLOv7's size by compressing the model's size, making the model lighter while yet ensuring that the detection accuracy is maintained.

Table 2: Network depth contrast of diverse YOLOv7 models.

(a) Model	**(b)Backbone: CSP1_X**			**(c)Neck: CSP2_X**				
	(d)First	**(e)Second**	**(f)Third**	**(g)First**	**(h)Second**	**(i)Third**	**(j)Fourth**	**(k)Fifth**
(l)YOLOv7s	(m)CSP1_1	(n)CSP1_	(o)CSP1-3	(p)CSP	(q)CSP2_1	(r)CSP-2_	(s)CSP-2_	(t)CSP-2_
(u)YOLOv7m	(v)SP1_2	(w)CSP1_	(x)CSP1-6	(y)CSP	(z)CSP2_2	(aa) CSP-2	(ab)SP-2	(ac) CSP-2
(ad)YOLOv7l	(ae) CSP1_3	(af)CSP1_	(ag) CSP1-9	(ah) CSP	(ai)CSP2_3	(aj)CSP-2	(ak) CSP-2	(al)CSP-2_
(am)YOLOv7x	(an) CSP1_4	(ao) CSP1_	(ap) CSP1-12	(aq) CSP	(ar)CSP2_4	(as)CSP-2	(at)CSP-2	(au) CSP-2

A convolution kernel pruning strategy is employed in order to regulate the Focus and CBL structure's number of associated convolution kernels, hence altering the network's overall width. It is seen in Table 3 that the YOLOV7 and YOLOV7 models have very different network widths. Table 4 shows that, when compared to other YOLOv7 models, our model selects a different number of convolution kernels in various module topologies. First and second CBL modules utilize 40 convolution kernels each. Third and fourth CBL modules use 80 convolution kernels each, while 320 and 640 kernels are employed respectively in the fourth and fifth CBL modules. By doing so, we may reduce the YOLOv7 network's width, enhance the pace at which objects are detected, and raise the average accuracy.

Table 3: Network width contrast of diverse YOLOv7 mockups.

Model	**Number of Convolution Kernels**				
	Focus	**1st CBL**	**2nd CBL**	**3rd CBL**	**4th CBL**
YOLOv7s	32	64	128	256	512
YOLOv7l	64	128	256	512	1024
YOLOv7x	80	160	320	640	1280
YOLOv7m	48	96	192	384	768
IYOLOv7	40	80	160	320	640

IV. RESULTS AND DISCUSSION

PyTorch was constructed using the Windows 10 operating system and the Python language on an HP Pavilion machine (Intel (R) Core, 8 GB memory; Python3.8 platform.

Stochastic gradient descent (SGD) is used to optimize network parameters in this study's IYOLOv7 network. In this example, we used the default values for all of the parameters except for batch size, which was set to 64 and all of the other

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *24 August 2023* ***Revised:*** *09 October 2023* ***Accepted:*** *18 October 2023*

parameters except for weight decay, learning rate, and iterations epochs, all of which were set to 0.001. There are 4000 samples in total, with 3000 in the training set and 1000 in the test set. Because batch size and learning rate have a momentous impact on model performance, we will fine-tune these parameters in order to produce a model that is both more efficient and more effective.

4.1. Performances metrics

To assess the effectiveness of the proposed approach, several measures are considered; including the confusion, accuracy, recall, precision and F1 score. The measurements are determined based on the confounding matrix used to measure performance of the proposed model.

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (1)$$

$$Recall = TP/(TP + FN) \quad (2)$$

$$Precision = TP/(TP + FP) \quad (3)$$

$$F1 - measure = 2 \times (Precision \times Recall)/(Precision + Recall) \quad (4)$$

Where, TP indicates the True Positive, FP represents the false positive, TN depicts the True Negative and TP provides the True Positive. Table 4 delivers the experimental examination of proposed IYOLOv7 model with other versions of YOLO network in terms of various metrics.

Table 4: Experimental Analysis of Proposed IYOLOv7 Network

Method	Accuracy	Recall	Precision	F1-measure
YOLOv7s	96.85	71.01	80.15	85.30
YOLOv7m	97.33	76.95	83.58	88.71
YOLOv7l	97.90	80.06	86.21	91.14
YOLOv7x	97.79	85.47	92.22	94.42
IYOLOv7	98.50	90.59	96.34	97.95

In the analysis of accuracy, the proposed model achieved 98.50%, where the other models of YOLO such as v5s, v5m, v51, v5x achieved nearly 96% to 97% only. Figure 6 shows the graphical representation of proposed model in terms of accuracy.

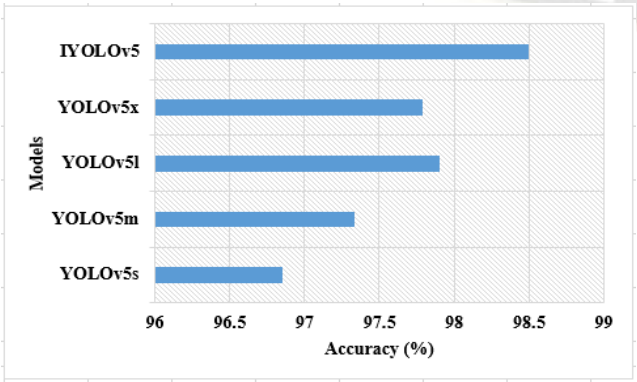

Figure 6: Accuracy Comparison

When comparing with all techniques, YOLOv7s achieved very low recall (i.e. 71.01%), where v5m achieved 76.95%, v51 achieved 80.06%, v5x achieved 85.47%. But, the improved layer of YOLOv7 achieved 90.59% of recall, which is shown in figure 7.

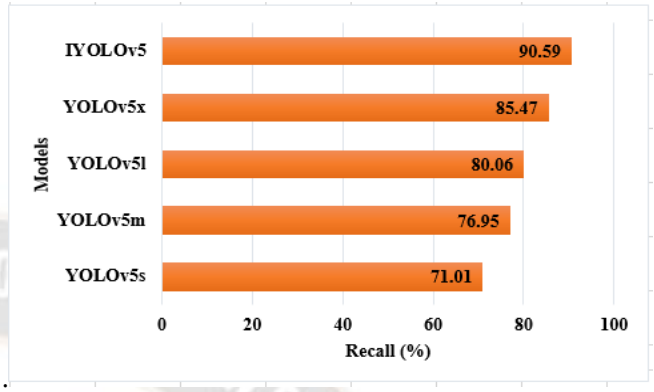

Figure 7: Recall Comparison of proposed improved model with other YOLO network

In the analysis of precision and F-measure, the proposed IYOLOv7 network achieved 96% to 97%, where YOLOv7m and v5s achieved nearly 84% to 86%. The YOLOv7x achieved 92.22% of precision and 94.42% of F1-measure, where graphical representation is provided in Figure 8 and 9.

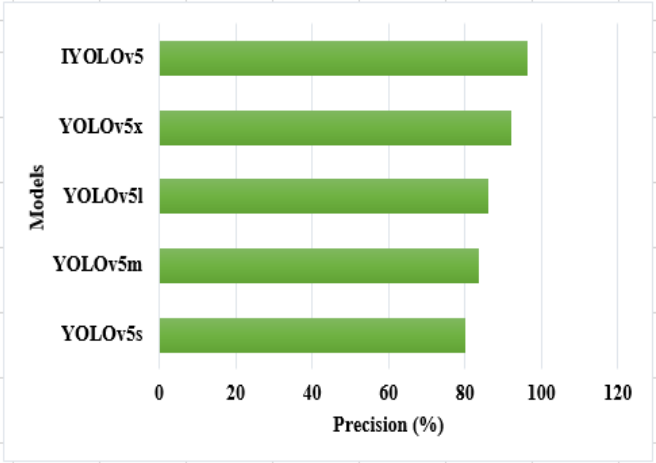

Figure 8: Precision Comparison

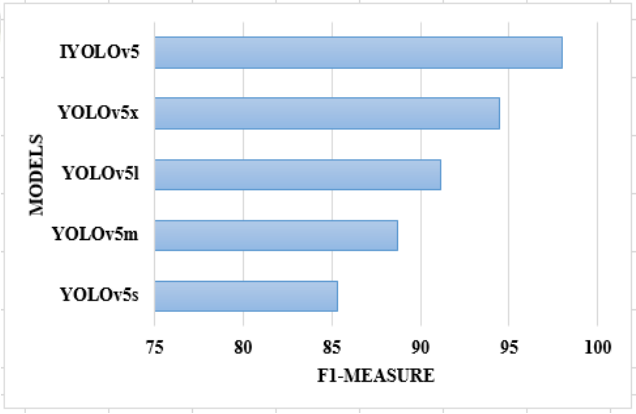

Figure 9: F1-Measure Comparison

In order to validate the training time and testing time of YOLO network, the results are shown in Table 5.

Table 5: Comparison of proposed model on Timing with Model Size

Method	Training time	Testing time	Model
YOLOv7s	22.532	1.32	16
YOLOv7m	19.453	1.45	14
YOLOv7l	20.130	1.13	43
YOLOv7x	21.631	1.25	95
IYOLOv7	18.145	0.58	12.5

When training the model, it takes more than one hour, it is because there are vast amount of data are used for learning the samples. For instance, YOLOv7s takes 22hrs, v5m takes 19hrs, v51 consumes 20hrs, v5x consumes 21hrs and proposed model consumed 18hrs. But, testing the data consumed less time, because the amount of data is less. For example, the other YOLO network takes nearly 1.45m to 1.13m, but the proposed model consumed 0.58m.

V. CONCLUSION

In this research, an IYOLOv7-based deep learning pupil tracking algorithm for wearable gaze trackers is presented. Using YOLO-based object recognition technology, the suggested pupil tracking approach accurately estimates and predicts the pupil's center in a visible-light mode. With the help of the created YOLO-based model, the accuracy and precision of pupil tracking can be up to 98 percent and 96 percent, respectively. The existing version of YOLO network achieved nearly 96% to 97% of accuracy, which shows clearly in the Section 4. In this design, the application distance between the user's head and the screen is fixed during the operation, and the direction of the head pose tries to keep fixed. In future works, a head movement compensation function will be added, and the proposed wearable gaze tracker will be more convenient and friendly for practical uses. To raise the high-precision recognition ability of the pupil location and tracking, the deep-learning model will be updated with optimization model for fine-tuning the learning rate or hidden neurons to fit the pupil position for different eye colors and eye textures.

REFERENCES

[1]. Duchowski AT. Eye tracking methodology: theory and practice. London: Springer, 2007.

[2]. Beltrán, J., García-Vázquez, M.S., Benois-Pineau, J., Gutierrez-Robledo, L.M. and Dartigues, J.F., 2018. Computational techniques for eye movements analysis towards supporting early diagnosis of Alzheimer's disease: a review. Computational and mathematical methods in medicine, 2018.

[3]. Morimoto CH and Mimica MRM. Eye gaze tracking techniques for interactive applications. Comput Vis Image Underst 2005; 98: 4–24.

[4]. Bates R, Istance H, Oosthuizen L, et al. D2.1 survey of defacto standards in eye tracking. Communication by Gaze Interaction (COGAIN). Deliverable 2.1, http://wiki.cogain.org/index.php/COGAIN_Reports (2005, accessed 25 April 2018).

[5]. Desanghere L and Marotta JJ. The influence of object shape and center of mass on grasp and gaze. Front Psychol 6. DOI: 10.3389/fpsyg.2015.01537

[6]. Halilaj, E., Rajagopal, A., Fiterau, M., Hicks, J.L., Hastie, T.J. and Delp, S.L., 2018. Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities. Journal of biomechanics, 81, pp.1-11.

[7]. Cognolato, M., Atzori, M. and Müller, H., 2018. Head-mounted eye gaze tracking devices: An overview of modern devices and recent advances. Journal of rehabilitation and assistive technologies engineering, 5, p.2055668318773991.

[8]. Castellini C and Sandini G. Learning when to grasp. In: Invited paper at Concept Learning for Embodied Agents, a workshop of the IEEE International Conference on Robotics and Automation (ICRA), Rome, Italy, 10–14 April 2007.

[9]. Novak D and Riener R. Enhancing patient freedom in rehabilitation robotics using gaze-based intention detection. In: IEEE 13th International Conference on Rehabilitation Robotics, ICORR, 2013, 24–26 June 2013, pp. 1–6. Seattle, WA, USA: IEEE.

[10]. Cognolato M, Graziani M, Giordaniello F, et al. Semiautomatic training of an object recognition system in scene camera data using gaze tracking and accelerometers. In: Liu M, Chen H and Vincze M (eds) Computer vision systems. ICVS 2017. Lecture notes in computer science, vol. 10528. Cham: Springer, pp.175–184.

[11]. Dos˘ en S, Cipriani C, Kostic´ M, et al. Cognitive vision system for control of dexterous prosthetic hands: experimental evaluation. J NeuroengRehabil 7. DOI: 10.1186/1743-0003-7-42.

[12]. Noronha B, Dziemian S, Zito GA, et al. 'Wink to grasp' – comparing eye, voice & EMG gesture control of grasp with soft-robotic gloves. In: International Conference on Rehabilitation Robotics, (ICORR) 2017, 17–20 July 2017, pp.1043–1048. London, UK: IEEE.

[13]. FU, W. and LIU, Y., 2020. Frontiers and progress in neuro-biomechanical ergogenic technology. Journal of Medical Biomechanics, pp.E649-E687.

[14]. Wang, S., Cui, L., Zhang, J., Lai, J., Zhang, D., Chen, K., Zheng, Y., Zhang, Z. and Jiang, Z.P., 2021, May. Balance control of a novel wheel-legged robot: Design and experiments. In 2021 IEEE International Conference on Robotics and Automation (ICRA) (pp. 6782-6788). IEEE.

[15]. Rubenstein LZ. Falls in older people: epidemiology, risk factors and strategies for prevention. Age Ageing 2006; 35: 37–41.

[16]. Majaranta P, Aoki H, Donegan M, et al. Gaze Interaction and Applications of Eye Tracking: Advances in Assistive Technologies. Hershey, PA: Information Science Reference – Imprints: IGI Publishing, 2011.

[17]. Goto S, Nakamura M and Sugi T. Development of meal assistance orthosis for disabled persons using EOG signal and dish image. Int J AdvMechatronSyst 2008; 1: 107–115.

[18]. Lin C-S, Ho C-W, Chen W-C, et al. Powered wheelchair controlled by eye-tracking system. Opt Appl 2006; 36: 401–412.

[19]. Ktena SI, Abbott W and Faisal AA. A virtual reality platform for safe evaluation and training of natural gaze-based wheelchair driving. In: 2015 7th International IEEE/EMBS Conference on Neural Engineering (NER), Montpellier, France, 22–24 April 2015, pp. 236–239. IEEE.

[20]. Blignaut P. Development of a gaze-controlled support system for a person in an advanced stage of multiple sclerosis: a case study. Univers Access InfSoc 2017; 16: 1003–1016.

[21]. Agarwal, A., JeevithaShree, D., Saluja, K. S., Sahay, A., Mounika, P., Sahu, A., Bhaumik, R., Rajendran, V. K., and Biswas, P. (2019b). Comparing two webcambased eye gaze trackersfor users with severe speech and motor impairment. In Chakrabarti, A., editor, Research into Design for a Connected World, pages 641–652, Singapore. Springer

[22]. Dong, X., Wang, H., Chen, Z., and Shi, B. E. (2015). Hybrid brain computer interface via bayesian integration of eeg and eye gaze. In 2015 7th International IEEE/EMBS Conference on Neural Engineering (NER), pages 150–153. IEEE

[23]. Li, W.Y.; Dong, Q.L.; Jia, H.; Zhao, S.; Wang, Y.C.; Xie, L.; Pan, Q.; Duan, F.; Liu, T.M. Training a Camera to Perform Long-Distance Eye Tracking by Another Eye-Tracker. IEEE Access 2019, 7, 155313–155324.

[24]. Rakhmatulina, I.;Duchowskim, A.T. Deep Neural Networks for Low-Cost Eye Tracking. In Proceedings of the 24th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems, Procedia Computer Science, Verona, Italy, 16–18 September 2020; pp. 685–694.

[25]. Brousseau, B.; Rose, J.; Eizenman, M. Hybrid Eye-Tracking on a Smartphone with CNN Feature Extraction and an Infrared 3D Model. Sensors 2020, 20, 543.

[26]. Kothari, R.S., Chaudhary, A.K., Bailey, R.J., Pelz, J.B. and Diaz, G.J., 2021. Ellseg: An ellipse segmentation framework for robust gaze tracking. IEEE Transactions on Visualization and Computer Graphics, 27(5), pp.2757-2767.

[27]. Madhusanka, B.G.D.A., Ramadass, S., Rajagopal, P. and Herath, H.M.K.K.M.B., 2022. Attention-aware recognition of activities of daily living based on eye gaze tracking. In Internet of Things for Human-Centered Design (pp. 155-179). Springer, Singapore.

[28]. Liu, M., Li, Y. and Liu, H., 2020. 3D gaze estimation for head-mounted eye tracking system with auto-calibration method. IEEE Access, 8, pp.104207-104215.

[29]. Guo, J., Liu, Y., Qiu, Q., Huang, J., Liu, C., Cao, Z. and Chen, Y., 2019. A Novel Robotic Guidance System With Eye-Gaze Tracking Control for Needle-Based Interventions. IEEE Transactions on Cognitive and Developmental Systems, 13(1), pp.179-188.

[30]. Bublea, A. and Căleanu, C.D., 2020, November. Deep Learning based Eye Gaze Tracking for Automotive Applications: An Auto-Keras Approach. In 2020 International Symposium on Electronics and Telecommunications (ISETC) (pp. 1-4). IEEE.

[31]. R. Valenti, N. Sebe, and T. Gevers, "Combining head pose and eye location information for gaze estimation," IEEE Trans. Image Process., vol. 21, no. 2, pp. 802–815, 2012, doi: 10.1109/TIP.2011.2162740.

[32]. P. M. Corcoran, F. Nanu, S. Petrescu, P. Bigioi, S. Member, and A. H. C. I. Eye-gaze, "Real-time eye gaze tracking for gaming design and consumer electronics systems," IEEE Trans. Consum. Electron., vol. 58, no. 2, pp. 347–355, 2012, doi: 0.1109/TCE.2012.6227433.

[33]. B. R. Pires, M. Hwangbo, M. Devyver, and T. Kanade, "Visible-spectrum gaze tracking for sports," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2013, pp. 1005–1010, doi: 10.1109/CVPRW.2013.146.