Scientific Journal of Informatics
Vol. 12, No. 3, August 2025
p-ISSN 2407-7658 https://journal.unnes.ac.id/journals/sji/index e-ISSN 2460-0040

SOCA-YOLO: Smart Optic with Coordinate Attention Model for Vision System-Based Eye Disease Detection

Rianto[1*], Vega Purwayoga[2], Aradea[3], Ali Astra Mikail[4], Irsalina Yumna[5]
[1, 2, 3, 4, 5]Department of Informatics, Siliwangi University, Indonesia

Abstract.
Purpose: The purpose of this research is to identify eye diseases using a modified YOLOv9. In particular, we modified YOLOv9 with the addition of Coordinate Attention (CA) for better eye disease detection performance, the use of Programmable Gradient Information (PGI), and Generalized Efficient Layer Aggregation Network (GELAN) for higher computational efficiency and accuracy.
Methods: This study consists of several stages, including the acquisition of eye disease data obtained from the Roboflow website, data annotation, image augmentation, modeling using a modified YOLOv9, and model evaluation.
Result: SOCA-YOLO model achieved an F1 score of 87,2% and mAP50 of 92,9%, outperforming YOLOv9-e by 1,7%. It also surpassed YOLOv6-L6 by 11,1%, YOLOv10-X by 0,8% in mAP50, and YOLOv8-X by 1,1% in recall, showcasing its superior detection accuracy and recall performance.
Novelty: This research contributes by introducing the SOCA-YOLO model in improving the performance of the YOLOv9 by modifying the addition of Coordinate Attention (CA) for better eye disease detection performance, alongside Programmable Gradient Information (PGI) and Generalized Efficient Layer Aggregation Network (GELAN) for better computational efficiency and accuracy.

Keywords: Coordinate attention, Eye disease, Object detection, YOLOv9
Received July 2025 / **Revised** August 2025 / **Accepted** September 2025

INTRODUCTION

Globally speaking, at least 2,2 billion people experience near or far vision impairment. Additionally, nearly 1 billion of these are preventable or untreated [1]. Among these 1 billion people, the paramount conditions causing distance vision impairment or blindness are cataracts (94 million), refractive errors (88,4 million), age-related macular degeneration (8 million), glaucoma (7,7 million) and diabetic retinopathy (3.9 million) [2]. Concerning near vision, rates of untreated impairment are estimated to be greater than 80% in western, eastern, and central sub-Saharan Africa. On the other hand, comparable rates in high-income regions (e.g., North America, Australasia, Western Europe, and Asia Pacific) were reported to be lower than 10% [3]. The obtained data from the results of the Rapid Assessment of Avoidable Blindness (RAAB) survey in 2014–2016 by the Ministry of Health disclosed that Indonesia indicated fairly serious blindness problems. It reached 30% and became the highest in Southeast Asia [4]. Further eye disease and the risk of vision loss can be prevented by detecting and treating eye problems earlier [5]. However, due to low awareness, a shortage of eye doctors, and high consultation costs, early eye checks are limited. Therefore, detecting eye diseases through automatic screening can be an alternative solution [6].

In the medical field, Convolutional Neural Network (CNN) architecture is commonly applied. As an example, identifying images related to medical matters can save time and costs because CNN has an effective feature extractor. This algorithm also signifies high accuracy compared to all other image prediction algorithms [7]. One type of CNN adopted for medical purposes is CataractNet. It was developed specifically to automatically detect cataracts in eye fundus images. It was designed with a customized Activation Function and Loss Function [8]. Other formulations of CNN (e.g., ResNet utilizing skip connections or shortcut connections) allow gradients to flow backward more efficiently through network layers (e.g., VGG16). In particular, it enables to adaptation of a CNN architecture consisting of 16 convolution layers and fully connected layers [9], [7].

*Corresponding author.
Email addresses: rianto@unsil.ac.id (Rianto)
DOI: 10.15294/sji.v12i3.29293

The Convolutional Neural Network (CNN) architecture continues to be developed and applied in several models. As an illustration, one of which is the You Only Look Once (YOLO) model accentuating real-time object detection with CNN as an integral part of its architecture [10]. A study conducted by [11] cultivated the EYE-YOLO model which is an improved variant of the Tiny YOLOv7 object detection model. EYE-YOLO highlighted the problem of eye disease recognition as a regression problem and bounding box detection by integrating multi-spatial pyramid pooling and focal-EIOU loss techniques to reinforce detection accuracy. In an inquiry conducted by [12], they focused on detecting the inner eye in the elderly by using YOLOv7 to train an eye detection model by performing transfer learning on the weights of the model trained on a larger multi-age dataset.

A study conducted by [13] applied the SENet attention module added to the YOLOv5 architecture. In this case, to improve detection performance characterized by small objects and high noise levels, researchers combined the proposed model with Feature Enhancement and Feature Denoising (FEFD). An investigative attempt by [14] developed the YOLOv4-LightC-CBAM model by reducing the width and depth of the backbone network. In particular, it transformed the neck network and improved the attention mechanism by including a Convolutional Block Attention Module (CBAM).

The YOLO model has speed in computing so it provides advantages in real-time object detection. Unfortunately, it suffers from lower accuracy compared to other object detection algorithms such as the two-stage method [10]. For this reason, there is still a void to enhance the speed and accuracy of the YOLO algorithm by considering or developing better algorithms [15].

Grounded in all the related work descriptions, there are still gaps open for future research, especially related to the results of fostering the CNN architecture through the YOLO model. In line with this, this study proposes an eye disease detection model called Smart Optic with Coordinate Attention (SOCA)-YOLO by cultivating the YOLOv9 model architecture [16] modified by adding Coordinate Attention (CA) [17] to the proposed model architecture.

The YOLOv9 model [16] has been selected because of its ability to increase speed and accuracy. As an example, YOLOv9 focuses on efficient data and parameter usage by utilizing the Programmable Gradient Information (PGI) concept and Generalized Efficient Layer Aggregation Network (GELAN) architecture proven to reduce the number of parameters and computational complexity without sacrificing object detection performance. Besides, YOLOv9 successfully reduces the number of parameters and computational complexity compared to previous versions of YOLO. Additionally, YOLOv9-E has 27% fewer calculations, and 16% fewer parameters, and it is 1,7% better than YOLOv8-X. Compared to YOLOv7 AF (Anchor-Free), YOLOv9-C has 42% fewer parameters and 22% less computation. However, it still reaches the same AP (Average Precision).

While YOLOv9 demonstrates a favorable trade-off between inference speed and detection accuracy, the model continues to exhibit limitations in identifying small-scale pathological features, such as microaneurysms or subtle retinal hemorrhages [18].

Modifications have been undertaken to the YOLOv9 architecture by adding an attention module such as CA [17] to improve model performance by maintaining attention channel position information and improve the performance results of the AP. Specifically, this aims to produce a spatial selective attention map. CA can compensate for the limitations of YOLOv9 through its ability to encode spatial information along both horizontal and vertical directions, thereby enhancing the representation of elongated patterns such as blood vessels while simultaneously preserving fine-grained details in small lesions [19]. In addition, CA factors channel attention into two one-dimensional feature coding processes that have been involved in combining features along two spatial directions.

To cultivate a generic formal framework for model development, the objective of this study is to develop a comprehensive framework for real-time eye disease detection and classification using computer vision techniques within the YOLO architecture. In particular, the system framework should be able to detect eye diseases through real-time eye screening and classify eye diseases wrapped in the YOLO architecture with the object detection method based on computer vision. The developed strategy is an eye disease detection model called SOCA-YOLO. The key contributions included integrating CA to improve feature extraction for more precise detection, utilizing Programmable Gradient Information (PGI) to optimize gradient flow

and accelerate training convergence, and incorporating the Generalized Efficient Layer Aggregation Network (GELAN) to achieve higher computational efficiency while reducing parameter usage. These innovations collectively optimize detection accuracy, enabling efficient automated eye screening and early diagnosis.

METHODS

Neural Networks frequently encountered challenges of information loss as input data underwent multiple layers of feature extraction and transformation resulting in loss of original information and reducing model accuracy. This problem could pose significant difficulties in eye detection tasks prioritizing accuracy in diagnosing eye diseases. To overcome such difficulties, the development of the SOCA-YOLO model was undertaken by developing the YOLOv9-e model architecture combined with the addition of CA [17] in the head section. YOLOv9 [16] was selected because it adapted Programmable Gradient Information (PGI) and Generalized Efficient Layer Aggregation Network (GELAN) to extract key features in data more effectively. Besides, it improved the accuracy of eye disease detection models by integrating miscellaneous processes, such as data annotation, pre-data processing, and data augmentation. The obtained model results went through a comprehensive analysis and evaluative stage through an evaluation matrix as an indicator for model evaluation. Figure 1 deciphers the model proposed in this research.

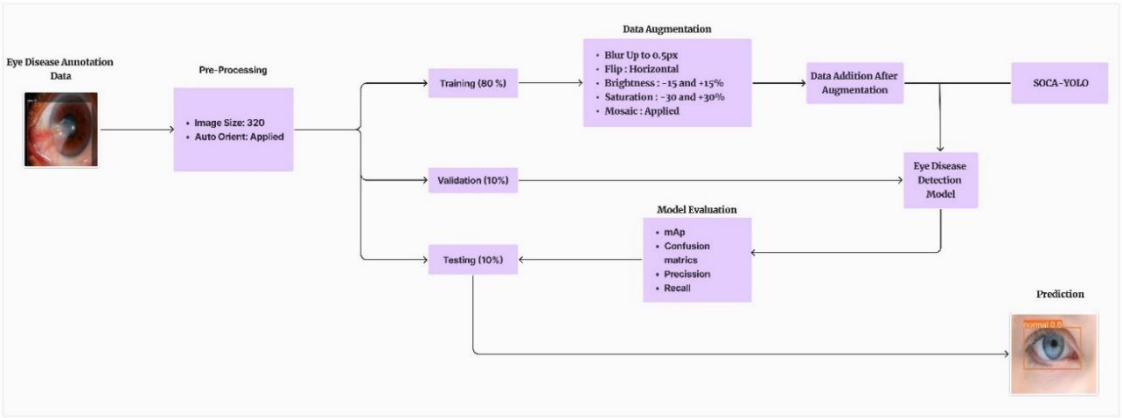


Figure 1. General architecture of the proposed model

Generally speaking, the proposed model architecture for eye disease detection utilized the YOLOv9 model. At the beginning of the model architecture flow, the annotated eye disease data was subsequently pre-processed with an image size of 320 x 320 pixels and automatic orientation. Then, it was divided into training (80%), validation (10%), and testing (10%) sets. Moreover, the data in the training set was subjected to data augmentation with rotation, horizontal flip, contrast adjustment (-15% to +15%), and saturation (-15% to +15%). The data expanded through augmentation and was adopted to train the YOLOv9-e model. The model was subsequently evaluated with mAP, F1-score, precision, and recall before being tested to produce a final prediction of eye disease. The architecture of the SOCA-YOLO model can be viewed in Figure 2.

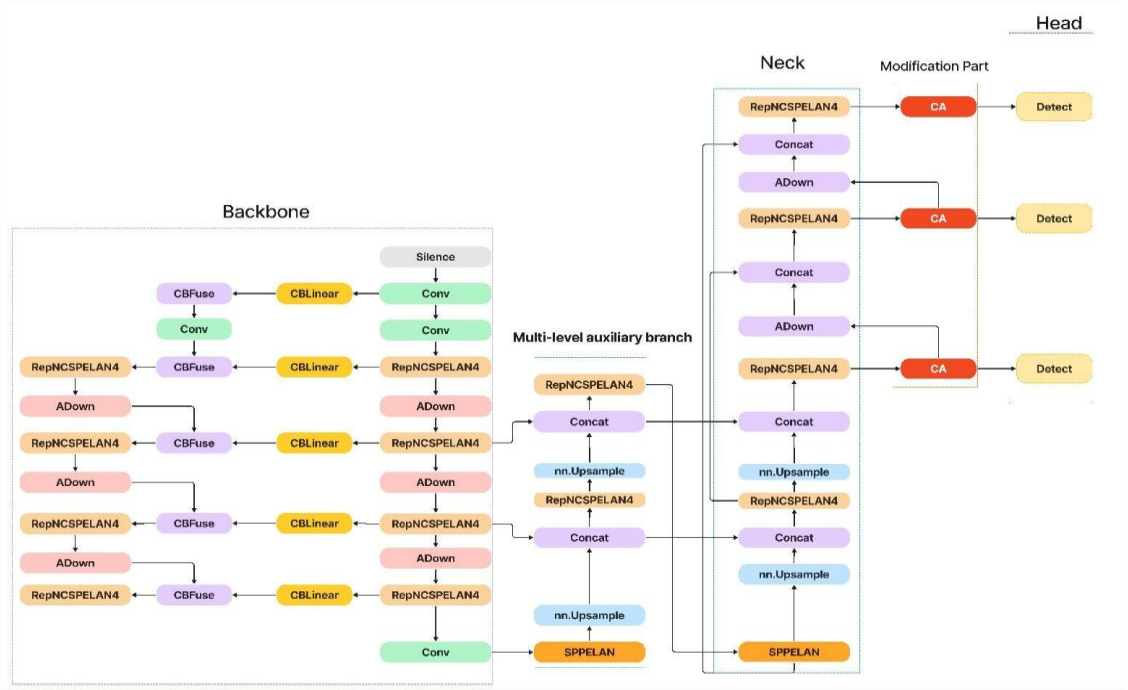


Figure 2. SOCA-YOLO model architecture

The SOCA-YOLO model architecture was a modification of the YOLOv9 [16] which is YOLOv9-e architecture by utilizing RepNCSP-ELAN4, RepNBottleneck, RepNCSP, and Adown blocks in the backbone to efficiently extract hierarchical features while maintaining spatial information. The PANet module in the neck was able to improve feature representation in more accurate object detection. The head in SOCA-YOLO functioned to optimize performance and ensure object detection results, including bounding box coordinates, class probabilities, and object scores. Modifications to the model architecture were performed on the neck by adding CA to the layer.

The process begins with the Backbone, which extracts features from the input data, typically using a pre-trained convolutional network to capture low-level and high-level information. The extracted features are then processed by the Multi-level Auxiliary Branch, which refines the features further, combines gradient information from the model's layers using specialized networks. By addressing the issue of information loss in deep supervision models, it guarantees that the model completely understands the data. The output from this auxiliary branch is then passed to the Neck, a component that performs additional feature fusion or transformations to bridge the gap between the Backbone and the final Head, which generates predictions. In this study, modifications to the model architecture were performed on the Neck by integrating an attention mechanism, namely CA. The placement of CA in the Neck enhances the model's ability to refine fused features by selectively emphasizing spatially significant regions while preserving positional information, which ultimately supports more accurate and context-aware predictions at the Head stage.

The stages of the eye disease detection method commenced with data acquisition and data collection of images of eye disease categorized into three classes, namely uveitis, cataract, and pterygium. After the data were collected, the next stage was to label the dataset based on the determined class. The augmentation stage was also executed to obtain varied image data with the final accurate eye disease detection model results. Model training was conducted through the YOLOv9 model as the base architecture of the SOCA-YOLO model for eye disease detection. Specifically, the flow of eye disease detection methods in this study can be noticed in Figure 3.

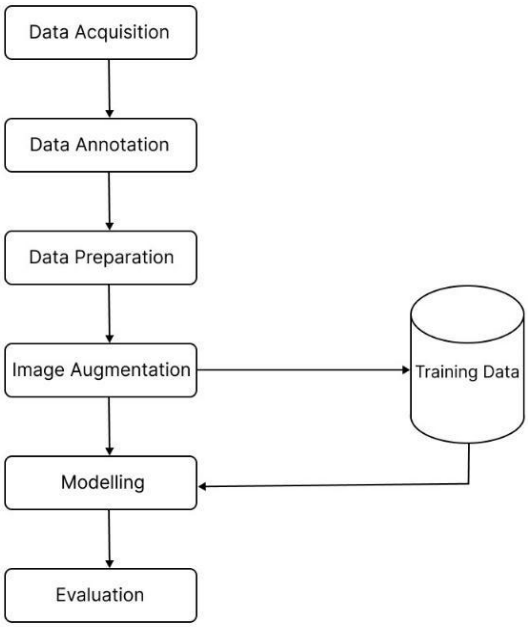


Figure 3. Flow of eye disease detection models

Data acquisition

This study applied datasets from the Roboflow website regarding human eye diseases (e.g., cataracts, pterygium, and uveitis). There were 2016 image datasets in creating eye disease detection models collected in the Eye Disease Classification dataset. Their eye disease image data were taken from the EYES dataset [20] before being acquired from the Eye Disease Classification dataset [21]. More specifically, the data were accessible on the Roboflow website. The dataset consisted of an image resolution size of 320 x 320 pixels. The dataset was classified into 4 classes. Each class represented a condition or type of disease in the human eye, namely normal, cataract, uveitis, and pterygium. The collected eye disease image dataset was adopted to train an eye disease detection model with SOCA-YOLO. Additionally, such a dataset was the main experimental data for this study.

Data annotation

Data annotation was performed on unlabeled images. The process of annotating eye disease image data was undertaken by assigning bounding boxes to the image based on the condition or disease with the existing class as a label. The annotated dataset was classified into 4 classes. Each class represented a condition or type of disease in the human eye, namely normal, cataract, uveitis, and pterygium. The labels assigned to the images were adjusted to the class or category of eye disease defined at the data acquisition stage. After conducting the data annotation, 479 data were identified to be in the normal class, 480 data were in the pterygium class, 550 data were in the uveitis class, and 479 data were in the cataract class.

Data preparation

The collected image dataset was pre-processed by adjusting the pixel size to 320x320 and applying auto-orient to the dataset. The input resolution of 320 × 320 pixels was selected as a trade-off between speed and accuracy. Although higher resolutions (e.g., 640 × 640) may improve detection of small objects, 320 × 320 enables up to three times faster inference with stable accuracy, which is suitable for real-time applications [22]. Conversely, the application of auto-orient (or rectangular training) enables the system to preserve the original aspect ratio by resizing the image accordingly and incorporating necessary padding, thereby maintaining the proportional integrity of elongated anatomical structures, such as blood vessels, as well as the morphology of the optic disc [23].

Image augmentation

Image augmentation was operationalized to prevent overfitting in the eye disease identification model. Nowadays, a widespread and well-accepted image data augmentation practice is geometric and color

augmentation [24], such as mirror images, cropping and translating images, changing image color palettes, color processing, and geometric transformations (rotation, resizing, and so on).

In this context, data augmentation algorithms (e.g., saturation change, blur addition, and horizontal flip) were adopted. By augmenting data with these miscellaneous methods, the information in the dataset was expanded with greater variations. Consequently, the model could learn from various situations that may occur in eye diseases. With this in mind, it helps prevent overfitting and boosts the abilities of the model to better and more accurately identify eye diseases on unknown test data.

Data splitting

The dataset was categorized into train, validation, and test datasets with a ratio of 80:10:10 applied in the modeling process for eye disease detection. In addition, this ratio is commonly used for object detection and the resulting model is relatively stable, this proportion of data can achieve an optimal balance between learning, tuning, and model evaluation [25], [26], [27].

Modelling

The YOLOv9 model in this study adapted the reference from Wang et al. [16]. In object detection, an approach was taken using Programmable Gradient Information (PGI) and Generalized Efficient Layer Aggregation Network (GELAN) integrated with the YOLOv9 model. To improve real-time inference capabilities, YOLOv9 applied a Generalized Efficient Layer Aggregation Network (GELAN) combining predominant features from CSPNet [28] and ELAN [29]. GELAN extended the capabilities of ELAN by allowing the nesting of any computational block outside the convolution layer. Hence, it allowed inference optimization to be applied at all layers.

The complex architecture of the YOLOv9 framework to improve model training and efficiency is expressed through Programmable Gradient Information. PGI prioritizes accurate and effective gradient backpropagation by integrating additional supervisory nodes designed to combat information congestion in deep neural networks. Three components are integrated to form PGI namely Main Branch, Auxiliary Reversible Branch, and Multi-Level Auxiliary Information which each play different yet interrelated roles in the model design.

Furthermore, in this study, the YOLOv9 baseline was modified by integrating CA modules into the architecture to enhance feature representation, especially for detecting eye diseases such as cataract, uveitis, and pterygium. Unlike conventional channel or spatial attention, Coordinate Attention decomposes spatial attention into two one-dimensional feature encoding processes along the horizontal and vertical directions. This enables the model to capture not only inter-channel dependencies but also precise positional information, which is crucial for localizing small and fine-grained abnormalities in eye regions. The integration of CA modules strengthens the backbone and neck by improving the ability to highlight relevant pathological features while suppressing redundant background noise. The overall architecture of the proposed model, termed SOCA-YOLO, is illustrated in Figure 2, where CA blocks are strategically embedded to refine feature extraction and improve detection accuracy without significantly increasing computational complexity.

Model evaluation

The results of the prediction model were evaluated through a confusion matrix. In the confusion matrix, there were several types, namely precision, recall, and F1 score. In this case, precision was the ratio of the value of true positive predictions compared to the total results with positive predictions. Recall was a comparison of the true positive predicted value with all true positive data. The F1 score represented a comparison of the average precision and recall values assigned. The accuracy and confusion matrix can be formed as in Equations (1), (2), and (3).

$$Precission = \frac{TP}{TP+FP} \tag{1}$$

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

$$F1\ score = 2\frac{(Recall\ X\ Precision)}{(Recall + Precision)} \tag{3}$$

True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) were indicators as a result of each detection describing whether the model was able to detect and classify objects correctly or not.

Mean average precision (mAp) served as an indicator of detection accuracy in target recognition tasks. mAp was obtained by calculating the average precision value for each category. Specifically, it was obtained through the integration of the Precision Recall (P-R) curve and averaging these values. The equation for mAP is outlined in Equation (4). AP refers to the average value of all precisions obtained under all possible recall levels where N was the number of classes or c.

$$mAP = \frac{1}{N} \sum_{C=1}^{N} AP_c \quad (4)$$

RESULTS AND DISCUSSIONS

Data preprocessing

This study adopted a dataset from the Roboflow site regarding several eye diseases in humans with a total of 2016 image datasets. In particular, consisted of 4 classes with different amounts of data. The resolution and size of the pixels in the dataset were 320 x 320 pixels for each image. Data annotation was performed on unlabeled images by situating bounding boxes on the images. The labels given to the images were adjusted to the class or category of eye disease defined at the data acquisition stage. After conducting the data annotation, 479 data were identified to be in the normal class, 480 data were in the pterygium class, 550 data were in the uveitis class, and 479 data were in the cataract class. At the image augmentation stage, augmentation techniques (e.g., horizontal flip, blur, mosaic, saturation, and brightness adjustment) were applied subsequently:

1. Flip: Horizontal
2. Saturation: Range -30% and +30%
3. Brightness: Range -15% and +15%
4. Blur: 0.5px
5. Mosaic: Applied

This process of image augmentation could enhance the number of training data from 1178 to 3534. In other words, it was able to expand the information in the training data with a greater variety of data. This could affect the ability of such a model to identify eye diseases more accurately. Also, it provided assorted situations that may occur in eye diseases.

Data splitting

At this stage, the entire dataset totaling 4372 was categorized into train, validation, and test with a ratio of 80:10:10. The application of YOLOv9 for eye disease detection as described in the methods section was undertaken in several schemes.

Modelling

The description of which is presented in Table 1. Each scheme is differentiated based on its architecture and optimization techniques.

Table 1. Scheme of model

Schema	Model Versions
1	SOCA-YOLO
2	YOLOv9-m + CA
3	YOLOv9-c + CA
4	YOLOv9-s + CA
5	YOLOv9-t + CA
6	YOLOv9 + CA
7	YOLOv9-e
8	YOLOv9-m
9	YOLOv9-c
10	YOLOv9-s
11	YOLOv9-t
12	YOLOv9

The subsequent step was to evaluate the model. In this section, a detailed explanation was assigned regarding the configuration in the experiment encompassing hyperparameter settings, test data sets used,

experimental configuration, and the evaluation process to measure the effectiveness of the SOCA-YOLO model in eye disease detection. To ensure the reliability of the proposed methodology, all experiments were operationalized with consistent hardware conditions. Experiments were performed on Google Collaboratory with Nvidia Tesla T4 16GB graphics specifications, 16GB RAM, Intel(R) Xeon(R) 2,30GHz CPU, Ubuntu 18.04 Operation System, and 112,6 GB storage memory. The parameters during model training to detect eye diseases covered epochs 50, batch size 28, image size 320, learning rate 0,01, anchor 5, and SGD optimizer. The descriptive configuration was aimed at providing a clear image of the experimental training settings and configurations.

Model evaluation

The determining parameters for determining the best model were precision, recall, F1 score, mAP50, and mAP50-95. The model performance comparison can be viewed in Table 2. This scheme can determine whether SOCA-YOLO is superior to YOLOv9 architecture or not. Referring to the results of the model performance comparison presented in Table 2, the best model was the SOCA-YOLO model. To illustrate, the SOCA-YOLO model was exceptional in precision, F1 score, and maAP50 parameters. However, the SOCA-YOLO model was less exceptional than the YOLOv9-e model in terms of Recall, and mAP50-95.

Table 2. Performance comparison with YOLOv9 model

Scheme	Model Versions	Precision (%)	Recall (%)	F1 score (%)	mAP50 (%)	mAP50-95 (%)
1	SOCA-YOLO	86,9	87,5	87,2	92,9	63,4
2	YOLOv9-m + CA	81,7	86,77	84,1	89,9	62,5
3	YOLOv9-c + CA	84,1	86,0	85,0	90,2	62,7
4	YOLOv9-s + CA	28,0	80,9	41,6	84,4	52,9
5	YOLOv9-t + CA	61,9	74,1	67,5	70,4	40,2
6	YOLOv9 + CA	80,4	80,7	80,5	87,5	57,0
7	YOLOv9-e	82,9	89,9	86,3	91,2	64,3
8	YOLOv9-m	83,0	87,4	85,1	90,5	61,9
9	YOLOv9-c	87,2	86,7	86,9	91,4	63,9
10	YOLOv9-s	76,0	76,5	76,2	83,6	51,8
11	YOLOv9-t	64,4	73,8	68,8	73,8	41,9
12	YOLOv9	83,4	74,5	78,7	84,2	55,0

The evaluation results indicate that the Recall and mAP50–95 of the model enhanced with CA do not surpass the performance of the YOLOv9-e baseline, this trade-off can still be justified. The relatively small decline in these metrics may be compensated by the model's improved ability to preserve spatial information, particularly for objects with elongated structures or subtle details. As reported in related studies, the CA mechanism enhances feature representation by effectively integrating positional and channel information. Therefore, despite the slight reduction in primary quantitative metrics, the incorporation of CA continues to contribute to detection robustness in medical imaging contexts, especially in cases where spatial accuracy is more critical than numerical performance alone.

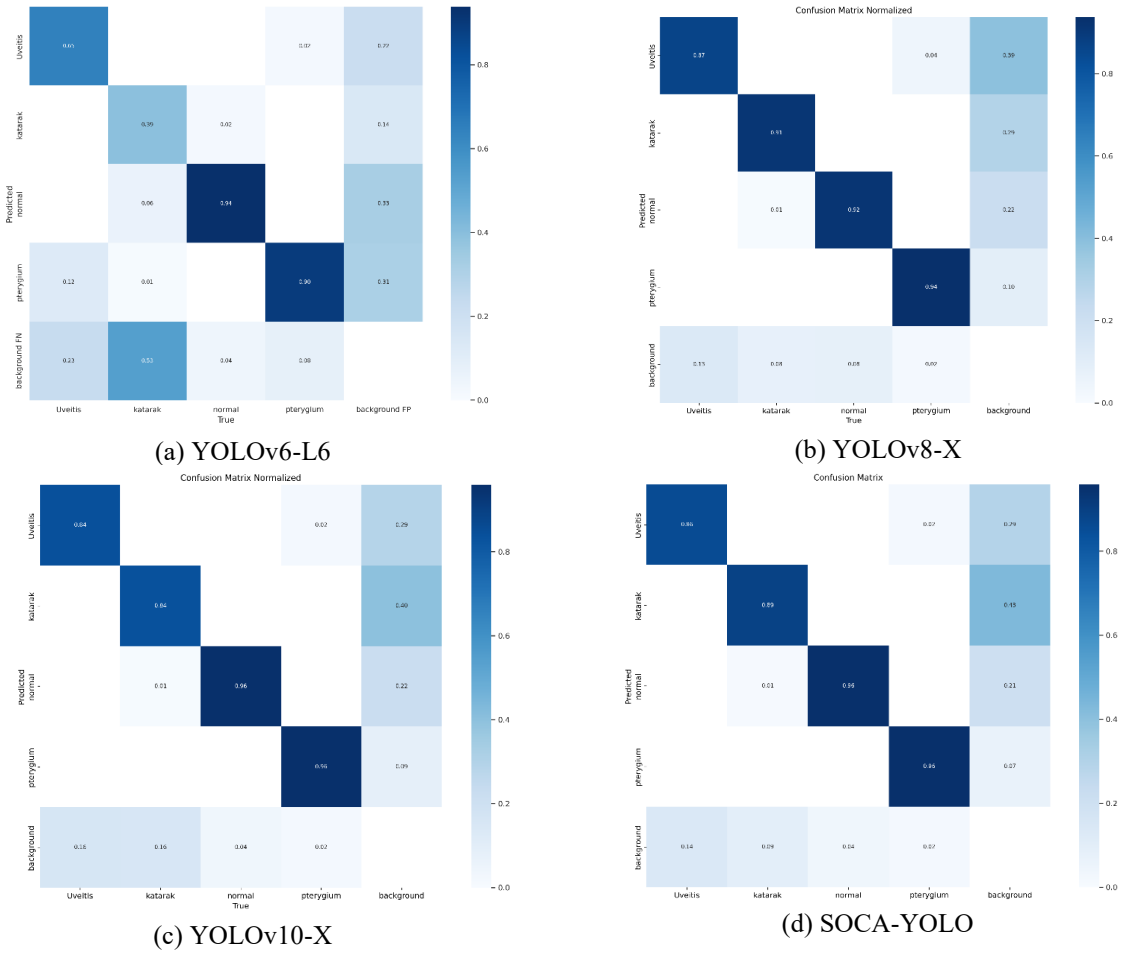
The model comparison process undertaken in this study was not only aimed at comparing the YOLOv9 model in Table 2 but also previous and subsequent versions. A comparison of the YOLOv9 model with previous and subsequent versions is presented in Table 3. Table 3 presents a comparison between the SOCA-YOLO model with the YOLOv8-X [30], YOLOv6-L6 [31], and YOLOv10-X [32] models. The results from Table 3 disclose that the YOLOv8-X model outperformed other models in the precision, F1 score, mAP50, and mAP50-95 parameters. On the other hand, the SOCA-YOLO model outperformed the recall parameters.

Table 3. Performance comparison with other YOLO models

Scheme	Model Versions	Precision (%)	Recall (%)	F1 score (%)	mAP50 (%)	mAP50-95 (%)
1	SOCA-YOLO	86,9	87,5	87,2	92,9	63,4
2	YOLOv8-X	91,8	86,4	89,0	93,6	70,0
3	YOLOv6-L6	81,8	67,0	73,7	81,8	51,8
4	YOLOv10-X	88,1	86,2	87,1	92,1	66,6

A more comprehensive visualization for evaluation and comparison of the SOCA-YOLO model with the YOLOv8-X, YOLOv6-L6, and YOLOv10-X models can be noticed in Figure 4 displays the confusion matrix. The confusion matrix aimed at indicating how to assess the performance of the model and compare it with other models. Specifically, a True Positive (TP) occurred when the model successfully detected an object in a real image. A False Positive (FP) took place when the network incorrectly detected an object in

an image. False Negatives (FN) occur when the network fails to detect an object in a fake image. These three elements are used to calculate the precision, recall and F1 Score used in Table 2 and Table 3.



(a) YOLOv6-L6 (b) YOLOv8-X

(c) YOLOv10-X (d) SOCA-YOLO

Figure 4. Comparison of Confusion Matrix for YOLO Models: (a) YOLOv6-L6, (b) YOLOv8-X, (c) YOLOv10-X, (d) SOCA-YOLO

Based on the results and experiments conducted, SOCA-YOLO showed superior performance implying that this model can provide more accurate and reliable detection, making it a valuable tool for early diagnosis in clinical settings. In addition, its effectiveness highlights the potential for further research on the use of real-time on portable medical devices, improving accessibility to eye care, especially in underserved areas. The robustness of this model also paves the way for expanding its use to detect other eye conditions or integrating it with broader AI-based diagnostic platforms to create comprehensive healthcare solutions.

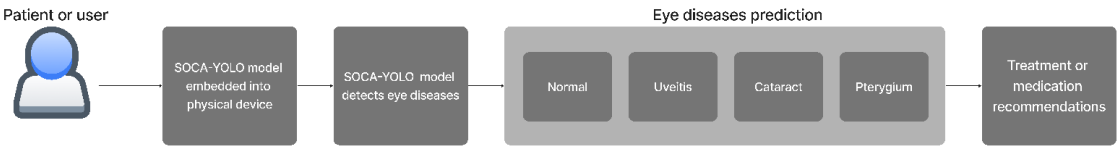


Figure 5. SOCA-YOLO Eye Disease Detection System

The flow of the eye disease detection system using SOCA-YOLO is shown in Figure 5 starts with the patient or user accessing a physical device that has been implanted with the SOCA-YOLO model. This device functions to detect eye diseases through captured eye images. After the SOCA-YOLO model

processes the image, the system will identify the user's eye condition and classify it into categories such as “Normal”, “Uveitis”, “Cataract”, or “Pterygium”. Based on the detection results, the system will provide recommendations regarding appropriate treatment or care. With this approach, patients can receive early detection and quick and precise medical recommendations automatically.

CONCLUSION

Grounded in the results of this study, the SOCA-YOLO method as a result of developing the YOLOv9-e model with modifications by adding Coordinate Attention (CA) can be applied for disease detection in humans. YOLOv9 utilizes the Programmable Gradient Information (PGI) concept and Generalized Efficient Layer Aggregation Network (GELAN) architecture proven to reduce the number of parameters and computational complexity without reducing object detection performance. The addition of CA plays an indispensable role in the SOCA-YOLO model facilitating the model to seek and recognize objects of interest more accurately. This study has successfully produced a model to detect assorted eye conditions (e.g., normal eyes, pterygium, cataracts, and uveitis) through the SOCA-YOLO model with an F1 score of 87,2% and mAP50 of 92,9%. Briefly stated, this study contributes to the development of object detection, especially the YOLOv9 model architecture through a series of experiments and evaluations. The investigative results can be further developed and applied in the health industry, notably in facilitating early detection of eye disease in patients and medical staff. In other words, the model enables to determination of the type of eye disease suffered by patients through automatic screening with the SOCA-YOLO model. However, it should be noted that this study still has room for improvement. To illustrate, the scope of this study is limited to the model performance testing stage and experiments with the CA module. Also, there are still investigative gaps related to the issue of employing other methods or algorithms to enhance the learning ability of the model while detecting objects more accurately. In future work, we plan to conduct clinical validation studies to assess the robustness and reliability of the proposed model in real-world diagnostic settings. This step is essential to ensure that the system not only performs well on curated datasets but also generalizes effectively to diverse clinical environments. Furthermore, we intend to explore the integration of the model into portable and resource-constrained devices, such as handheld or wearable diagnostic tools, to facilitate point-of-care applications. Such integration would enhance the accessibility and usability of the system, particularly in regions with limited medical infrastructure.

REFERENCES

[1] World Health Organization, “Blindness and Vision Impairment,” Sep. 10, 2023. Accessed: Aug. 29, 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment

[2] R. R. A. Bourne *et al.*, “Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: The Right to Sight: An analysis for the Global Burden of Disease Study,” *Lancet Glob Health*, vol. 9, no. 2, 2021, doi: 10.1016/S2214-109X(20)30489-7.

[3] T. R. Fricke *et al.*, “Global Prevalence of Presbyopia and Vision Impairment from Uncorrected Presbyopia: Systematic Review, Meta-analysis, and Modelling,” *Ophthalmology*, vol. 125, no. 10, 2018, doi: 10.1016/j.ophtha.2018.04.013.

[4] KEMENTRIAN KESEHATAN REPUBLIK INDONESIA, “PETA JALAN PENANGGULANGAN GANGGUAN PENGLIHATAN DI INDONESIA TAHUN 2017 – 2030.” Accessed: Aug. 29, 2024. [Online]. Available: https://p2ptm.kemkes.go.id/uploads/VHcrbkVobjRzUDN3UCs4eUJ0dVBndz09/2018/08/Buku_Peta_Jalan_Penanggulangan_Gangguan_Penglihatan_di_Indonesia_tahun_2017_2030.pdf

[5] M. N. I. Muhlashin and A. Stefanie, “Klasifikasi Penyakit Mata Berdasarkan Citra Fundus Menggunakan YOLO V8,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 2, 2023, doi: 10.36040/jati.v7i2.6927.

[6] S. Chelaramani, M. Gupta, V. Agarwal, P. Gupta, and R. Habash, “Multi-task knowledge distillation for eye disease prediction,” in *Proceedings - 2021 IEEE Winter Conference on Applications of Computer Vision, WACV 2021*, 2021. doi: 10.1109/WACV48630.2021.00403.

[7] A. O. Asia *et al.*, “Detection of Diabetic Retinopathy in Retinal Fundus Images Using CNN Classification Models,” *Electronics (Switzerland)*, vol. 11, no. 17, 2022, doi: 10.3390/electronics11172740.

[8] M. S. Junayed, M. B. Islam, A. Sadeghzadeh, and S. Rahman, “CataractNet: An automated cataract detection system using deep learning for fundus images,” *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3112938.

[9] R. Amin, A. Ahmed, S. Shabih Ul Hasan, and H. Akbar, "Multiple Eye Disease Detection Using Deep Learning," *Foundation University Journal of Engineering and Applied Sciences
(HEC Recognized Y Category , ISSN 2706-7351)*, vol. 3, no. 2, 2023, doi: 10.33897/fujeas.v3i2.689.

[10] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016. doi: 10.1109/CVPR.2016.91.

[11] A. Kumar and R. Dhanalakshmi, "EYE-YOLO: a multi-spatial pyramid pooling and Focal-EIOU loss inspired tiny YOLOv7 for fundus eye disease detection," *International Journal of Intelligent Computing and Cybernetics*, vol. 17, no. 3, pp. 503–522, Jul. 2024, doi: 10.1108/IJICC-02-2024-0077.

[12] M. Ghourabi, F. Mourad-Chehade, and A. Chkeir, "Eye Recognition by YOLO for Inner Canthus Temperature Detection in the Elderly Using a Transfer Learning Approach," *Sensors*, vol. 23, no. 4, 2023, doi: 10.3390/s23041851.

[13] Y. Zhang, Y. Qiu, and H. Bai, "FEFD-YOLOV5: A Helmet Detection Algorithm Combined with Feature Enhancement and Feature Denoising," *Electronics (Switzerland)*, vol. 12, no. 13, 2023, doi: 10.3390/electronics12132902.

[14] Z. Cao and R. Yuan, "Real-Time Detection of Mango Based on Improved YOLOv4," *Electronics (Switzerland)*, vol. 11, no. 23, 2022, doi: 10.3390/electronics11233853.

[15] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," Apr. 2018, [Online]. Available: ArXivID: 1804.02767

[16] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information," Feb. 2024, [Online]. Available: ArXivID: 2402.13616

[17] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2021. doi: 10.1109/CVPR46437.2021.01350.

[18] T. Wang *et al.*, "Intelligent Diagnosis of Multiple Peripheral Retinal Lesions in Ultra-widefield Fundus Images Based on Deep Learning," *Ophthalmol Ther*, vol. 12, no. 2, 2023, doi: 10.1007/s40123-023-00651-x.

[19] Y. Nie, J. Hou, X. Han, and M. Nießner, "RFD-Net: Point Scene Understanding by Semantic Instance Reconstruction," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2021. doi: 10.1109/CVPR46437.2021.00458.

[20] Roboflow, "Eyes." Accessed: Sep. 08, 2024. [Online]. Available: https://universe.roboflow.com/yi-2jxqr/eyes-kjjbv

[21] Roboflow, "Klasifikasi Penyakit Mata." Accessed: Sep. 05, 2024. [Online]. Available: https://universe.roboflow.com/hibahpenelitian/klasifikasi-penyakit-mata/dataset/7

[22] J. Chai, H. Zeng, A. Li, and E. W. T. Ngai, "Deep learning in computer vision: A critical review of emerging techniques and application scenarios," *Machine Learning with Applications*, vol. 6, 2021, doi: 10.1016/j.mlwa.2021.100134.

[23] H. Qiu, H. Li, Q. Wu, F. Meng, K. N. Ngan, and H. Shi, "A2RMNet: Adaptively Aspect Ratio Multi-Scale Network for object detection in remote sensing images," *Remote Sens (Basel)*, vol. 11, no. 13, 2019, doi: 10.3390/rs11131594.

[24] D. Sun and F. Dornaika, "Data augmentation for deep visual recognition using superpixel based pairwise image fusion," *Information Fusion*, vol. 107, 2024, doi: 10.1016/j.inffus.2024.102308.

[25] I. Grijalva, B. J. Spiesman, and B. McCornack, "Computer vision model for sorghum aphid detection using deep learning," *J Agric Food Res*, vol. 13, 2023, doi: 10.1016/j.jafr.2023.100652.

[26] S. Nuankaew, N. Boonyuen, K. Thumanu, and N. Pornputtapong, "Development of a machine learning model for systematics of Aspergillus section Nigri using synchrotron radiation-based fourier transform infrared spectroscopy," *Heliyon*, vol. 10, no. 5, 2024, doi: 10.1016/j.heliyon.2024.e26812.

[27] M. R. Shoaib *et al.*, "Deep learning innovations in diagnosing diabetic retinopathy: The potential of transfer learning and the DiaCNN model," *Comput Biol Med*, vol. 169, 2024, doi: 10.1016/j.compbiomed.2023.107834.

[28] C. Y. Wang, H. Y. Mark Liao, Y. H. Wu, P. Y. Chen, J. W. Hsieh, and I. H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2020. doi: 10.1109/CVPRW50498.2020.00203.

[29] C. Y. Wang, H. Y. M. Liao, and I. H. Yeh, "Designing Network Design Strategies Through Gradient Path Analysis," *Journal of Information Science and Engineering*, vol. 39, no. 2, 2023, doi: 10.6688/JISE.202307_39(4).0016.
[30] G. Jocher, A. Chaurasia, and J. Qiu, "YOLOv8 by Ultralytics." Accessed: Feb. 02, 2024. [Online]. Available: https://github.com/ultralytics/ultralytics
[31] C. Li *et al.*, "YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications." [Online]. Available: https://github.com/meituan/YOLOv6.
[32] W. A *et al.*, "YOLOv10: Real-Time End-to-End Object Detection", Accessed: Sep. 30, 2025. [Online]. Available: http://arxiv.org/abs/2405.14458