Article

Not peer-reviewed version

Traffic Light Recognition Assistant for Color Vision Deficiency Using YOLO with Multilingual Audio Feedback

Yinyuan Ma , Fathan Arifah * , Qonita Afifah , Liko Bun , Kangfu Zhang , Minan Tang

Posted Date: 8 January 2026

doi: 10.20944/preprints202601.0608.v1

Keywords: color vision deficiency; traffic light; YOLOv12; computer vision; audio feedback

Article

Traffic Light Recognition Assistant for Color Vision Deficiency Using YOLO with Multilingual Audio Feedback

Ma Yinyuan [1], Fathan Arifah [1,*], Qonita Afifah [2], Liko Bun [3], Zhang Kangfu [1] and Minan Tang [4]

[1] School of Mechanical and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China
[2] Fakultas Teknologi Industri, Universitas Ahmad Dahlan (UAD), Special Region of Yogyakarta 55166, Indonesia
[3] College of International Student Education, Chongqing Medical University, Chongqing 400016, China
[4] School of Automation and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou, 730070, China
* Correspondence: 2023g28004@stu.lzjtu.edu.cn

Abstract

Drivers with color vision deficiency (CVD) often face difficulty recognizing traffic light colors at intersections, putting at risk their safety and independence while driving in city environments. This study presents the development of an assistive prototype designed with Python and a PyQt5 graphical user interface. The system applies a YOLOv12 model, a Convolutional Neural Network-based object identification method that uses the OpenCV Python library that has been trained and evaluated on a comprehensive dataset consisting of various conditions, such as daytime and nighttime circumstances, clear and rainy weather, and traffic density, to recognize traffic light signals as red, yellow, and green. The detection result of traffic light color from a car webcam is delivered to users with offline audio feedback available in Indonesian, Mandarin, and English. During testing, we found a maximum confidence of 0.95 across eight challenging scenarios. The system aims to improve driving safety for individuals with color vision deficiency, offering an additional assistive device rather than replacing standard driving regulations.

Keywords: color vision deficiency; traffic light; YOLOv12; computer vision; audio feedback

1. Introduction

Traffic safety in urban environments strongly depends on drivers' ability to interpret traffic signals accurately. However, for people with Color Vision Deficiency (CVD), commonly known as color blindness [1]. CVD affects approximately 8% of the male population and 0.5% of the female population worldwide [2]. In various regions, such as Nepal, the prevalence of CVD is often an "iceberg phenomenon" where many individuals are unaware of their visual limitations until they encounter failure in critical tasks or undergo formal screening [3]. Studies of medical students in India also confirm that many future professionals are unaware of their own CVD, which could potentially affect their future clinical performance [4].

Clinically, the main difficulty lies in the failure of cone cells in the retina to distinguish between certain wavelengths of light, particularly between the red and green spectrums. The impact is not limited to comfort, but also to the ability to participate safely in modern mobility [5]. Research shows that people with CVD require longer reaction times to identify changes in traffic lights compared to those with normal trichromatic vision [6]. This is exacerbated by the fact that overall visual impairment correlates directly with an increase in traffic accidents, especially in densely populated urban areas with high levels of visual pollution [7]. Beyond navigation, poor color discrimination also hinders an individual's ability to interpret scientific data or spatial information, necessitating inclusive color standardization [8,9].

Regulations regarding driver's license ownership are highly inconsistent globally. In the ASEAN region, countries such as Indonesia, Cambodia, and Thailand enforce very strict rules through the Ishihara test [10]. In many cases, this test acts as an absolute barrier to obtaining a driving license, despite ongoing academic debate regarding the direct relationship between CVD and fatal traffic accidents [11]. Research in the European context suggests that color vision impairment does not pose a significant threat to traffic safety, leading many countries in Europe, Canada, and Oceania to eliminate specific color vision standards for private drivers [12]. In the United Kingdom, for instance, the government explicitly states that CVD is not a notifiable condition, and drivers are permitted to drive without medical consultation as long as they meet standard visual acuity and field requirements [13]. This inconsistency creates inequality in access to personal mobility. On the other hand, physical infrastructure often fails to provide adequate assistance. Although traffic lights have a standard position, design variations, environmental influences such as fog, or glare from the sun mean that position indicators alone are not a reliable solution. This encourages the need for technology-based intervention strategies, such as digital image re-coloring techniques [14].

Due to the limitations of current traffic light detection systems related to the safety risks for transportation users, especially in developing countries, this study aims to develop a detection system that is reliable by using the YOLOv12 model. The main objective of this study is to improve driving safety by improving upon simple color recognition and focusing on the structural and spatial context of traffic signals. By training the model to recognize the entire traffic light frame, the system can interpret instructions based on active lights in vertical and horizontal configurations.

The main contributions of this research are as follows:

- Spatial Position Detection: We implement a detection framework using full-frame annotations. This provides high confidence by making use of the position of the lights within the frame, which effectively reduces the risk of color misclassification for drivers with CVD.
- Robust Real-World Dataset: We introduce datasets recorded in urban environments, including lighting conditions (day and night), weather (clear and rainy), and traffic density (normal and crowded) scenarios.
- Assistive Technology: The system is an additional tool that provides offline multilingual audio feedback (Indonesian, Mandarin, and English) to improve reaction times for drivers.

This research proposes reliable technological intervention to assist the color-blind community in achieving safer mobility.

2. Related Works

This section reviews recent studies related to traffic light recognition and assistive technologies for users with color vision deficiency, focusing on CNN-based traffic light recognition, visual sensing systems in vehicles, voice feedback integration for smart transportation, and navigation assistance.

2.1. Deep Learning-Based Traffic Light Recognition

The Deep Learning Revolution in Computer Vision The transformation of assistive solutions is now shifting to Deep Learning (DL)-based intelligent systems. The theoretical basis of DL enables computers to automatically learn complex features from image data [15]. A major breakthrough in this field is the introduction of the You Only Look Once (YOLO) algorithm, which introduced integrated real-time object detection patterns [16]. The ability of YOLO to process video frames at high speeds makes it useful in mobile applications. However, for mobile devices is limited computational resources the main challenge. To overcome this issue, architectures such as MobileNetV2 enable complex models to run efficiently on mobile devices [17].

In the context of transportation, the implementation of Convolutional Neural Networks (CNN) has reached an extraordinary level of proficiency in recognizing traffic lights [18]. Recent surveys show that CNN is capable of overcoming detection challenges in dynamic real-world environments [19,20]. One of the technical obstacles that often arises is motion blur caused by fast-moving vehicles;

however, image restoration methods can now solve this problem [21]. Additionally, the integration of priority maps enables the system to verify the relevance of traffic lights on specific routes [22]. This technology has even begun to be applied to Raspberry Pi-based systems to distinguish between traffic lights and vehicle taillights, which can often be confusing for people with CVD at night [23].

2.2. Assistive Technologies and Voice Feedback Systems

Assistive Technology and Voice Feedback Integration Today's assistive technology no longer only displays information visually, but also through voice feedback integration (voice alerts) [24,25]. Python-based object detection systems with voice output have proven to be very effective in assisting human-machine interaction for those with visual impairments [26]. The use of libraries such as OpenCV and HSV color space manipulation also facilitates precise object color identification in Android applications [27–29]. For blind pedestrian navigation, models such as TFP-YOLO [30] have been optimized to detect small obstacles on the road with an attention mechanism. YOLOv12 introduces an attention-based architecture that demonstrates highly accurate small object detection performance and has been successfully tested in extreme conditions, such as underwater object detection [31,32]. Research over the past five years has shown a significant increase in accuracy in the development of color object detection systems using machine learning, which is capable of displaying color names on specific pixels to assist in daily activities [33]. The Convolutional Neural Network (CNN) method has also been developed for real-time traffic light color recognition with high accuracy [34]. In addition, the FlashLightNet model is capable of detecting light status, both static and flashing, end-to-end [35]. Improvements are also being made to YOLOv3 [36] and YOLOv4 [37] to optimize traffic sign detection, thereby improving road safety through smarter CNN-based traffic sign classification [38]. Several innovations have developed systems that combine the use of YOLO with voice feedback to assist people with visual impairments [39,40]. In addition, voice guidance generated through text-to-speech (TTS) allows users to find objects in a space without the need for assistance from others [41].

2.3. Research Rationale

Despite significant advances in CNN-based traffic light recognition systems and voice feedback assistance systems, there are still several critical gaps in the current literature. First, many existing models [18,19,34] prioritize color identification, which poses a safety risk in scenarios where color perception is impaired by weather, glare, or visual impairments such as CVD. Second, although advanced models such as YOLOv4 to YOLOv12 have been applied to general object detection [32,36], research on training these models to recognize the complete structure of traffic lights in order to utilize spatial-positional context is still limited.

Furthermore, most studies focus on standard vertical signals or specific ideal lighting conditions, often neglecting the need for various orientation (vertical and horizontal) of diverse urban infrastructure. Although audio feedback has been explored for the visually impaired [39,40], its integration into a fully optimized real-time multilingual offline system for CVD drivers remains understudied. This study fills these gaps by developing a YOLOv12-based application that leverages spatial recognition and diverse real-world datasets (including daytime, nighttime, clear weather, rain, and traffic density) to provide safer and more reliable assistive tools for the CVD community.

3. Materials and Methods

The system in this study is divided into two main phases. First, focusing on training a customized YOLOv12 object detection model by labeling traffic light images according to their color classification. The images cover lighting conditions, weather, and traffic density. The main objective is to optimize model parameters and obtain weights that can classify traffic lights as red, yellow, green with high accuracy and confidence in various scenarios.

The second phase is developing application prototype. The best training weights are used to process video input taken directly from an attached vehicle webcam. The system detects traffic light conditions and immediately generates an audio feedback to the user in the selected language. This step enables quick assistance in the form of audio and visual color labels, as shown in Figure 1.

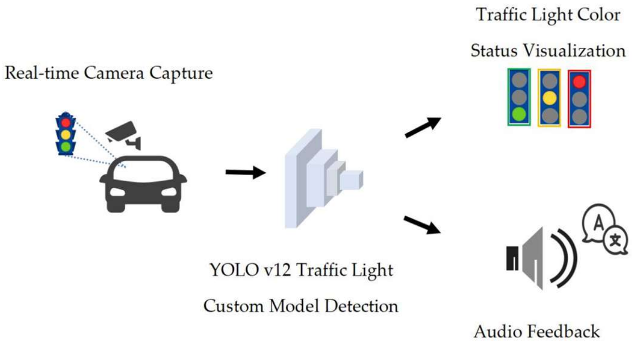


Figure 1. Proposed system.

2.1. Dataset Preparation

The dataset was collected using camera video recordings taken at locations in various urban areas in Indonesia and China. The dataset is organized in the following manner:

- Videos were converted into image frames using a video-to-frame extraction process;
- The extracted images were annotated according to predefined traffic light color categories;
- Each category contains between 200 and 600 images, for a total of 1,000 annotated images;
- Annotation was performed manually using Label Studio, with bounding boxes marking the entire traffic light;
- Annotated datasets were extracted in the standard YOLO format, including images and labels;
- The dataset was then randomly divided into 80% training data and 20% validation data.

2.2. Model Custom Training

The training runs on a system equipped with an Intel Core i7-3770 processor operating at 3.40GHz. The model was trained for 100 epochs, with approximately 125 batches per epoch. The number of batches per epoch is determined by dividing the training dataset size by a batch size of 8. Additional training settings include an input image dimension of 640 × 640 pixels, an initial learning rate of 0.01 with linear decay, and a Stochastic Gradient Descent (SGD) optimizer with momentum of 0.937. During the training process, the box loss, classification loss, and distribution focus loss gradually decreased from the initial epoch to the 100th epoch, as shown in Figure 2.

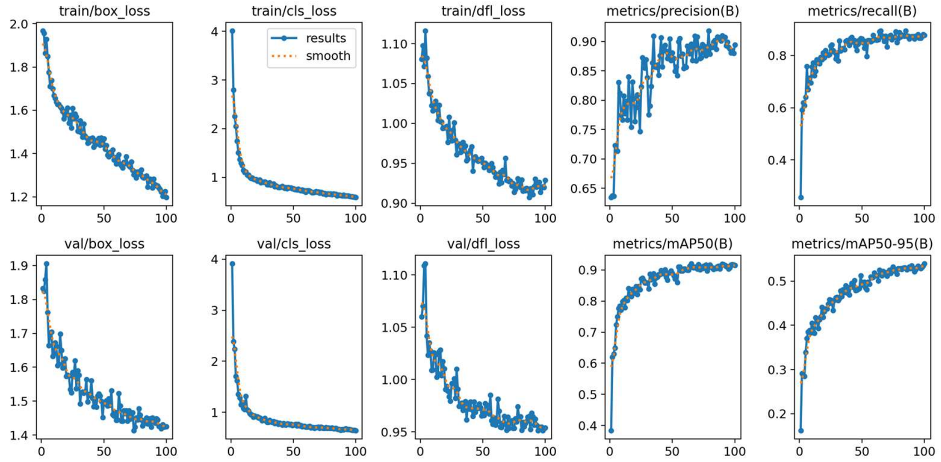


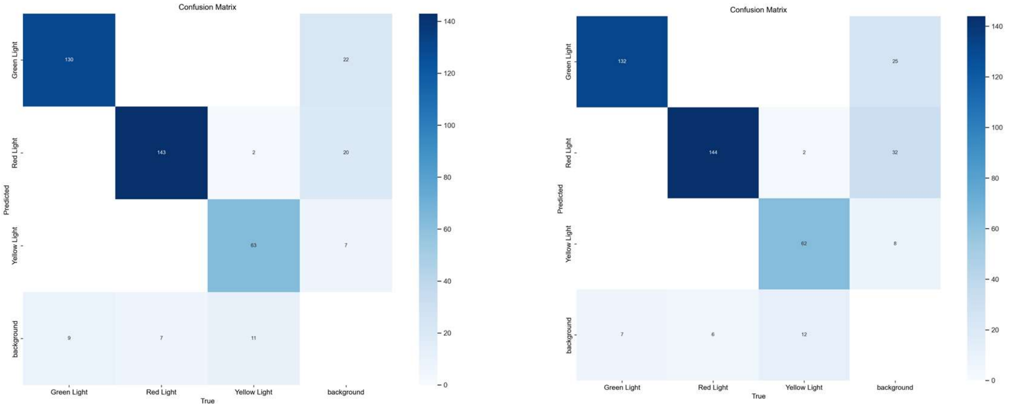
Figure 2. YOLO12 Model Custom Pre-trained Results.

Evaluation results model training demonstrated a mean Average Precision (mAP@0.5) between 0.92 and 0.95 at completion of training, whereas mAP@0.5:0.95 varied from 0.53 to 0.54. The precision at the 100th epoch attained a range of 0.89 to 0.91, while the recall achieved a range of 0.87 to 0.88. Other various of YOLO model including YOLO 8n, YOLO 10n, YOLO 11n, and YOLO 12n was also trained to ascertain optimal performance on the identical dataset. The training outcomes for each model are displayed in Table 1. YOLO 11n attained the best precision at 0.927 and recall at 0.895, although YOLO 12n exhibited the highest mAP@0.95 at 0.539. Simultaneously, YOLO 10n demonstrated the lowest performance among the four models, achieving a recall of 0.840.

Table 1. Model Pre-trained Comparison.

Model	Precision	Recall	mAP@50	mAP@95
YOLO 8n	0.916	0.895	0.923	0.537
YOLO 10n	0.892	0.840	0.892	0.524
YOLO 11n	0.927	0.895	0.913	0.532
YOLO 12n	0.918	0.895	0.920	0.539

YOLO 11n and YOLO 12n are suitable for use in detection systems, both showing a balance between precision, recall, and mAP. The selection of YOLOv12 as the final model is based on the confusion matrix analysis shown in Figure 3. YOLO 11n produces 25 false negatives, while YOLO 12n produces 27 false negatives, indicating that YOLO 11n misses fewer objects that actually exist. YOLO 11n also returns a higher number of false positives in the background, with 65 backgrounds incorrectly detected as objects compared to YOLO 12n's 49. YOLOv12 offers consistency in distinguishing between valid objects and background areas, including minority categories such as yellow lights. Given the traffic application requirements, we prioritize a balance between sensitivity and stability. Although YOLOv12 has a marginally higher false negative rate (a difference of only 2 instances), it significantly reduces false positives (by 16 instances). Minimizing false positives is crucial to prevent driver distraction and 'alarm fatigue,' ensuring the driver continues to trust and use the assistive system.



(**a**) (**b**)

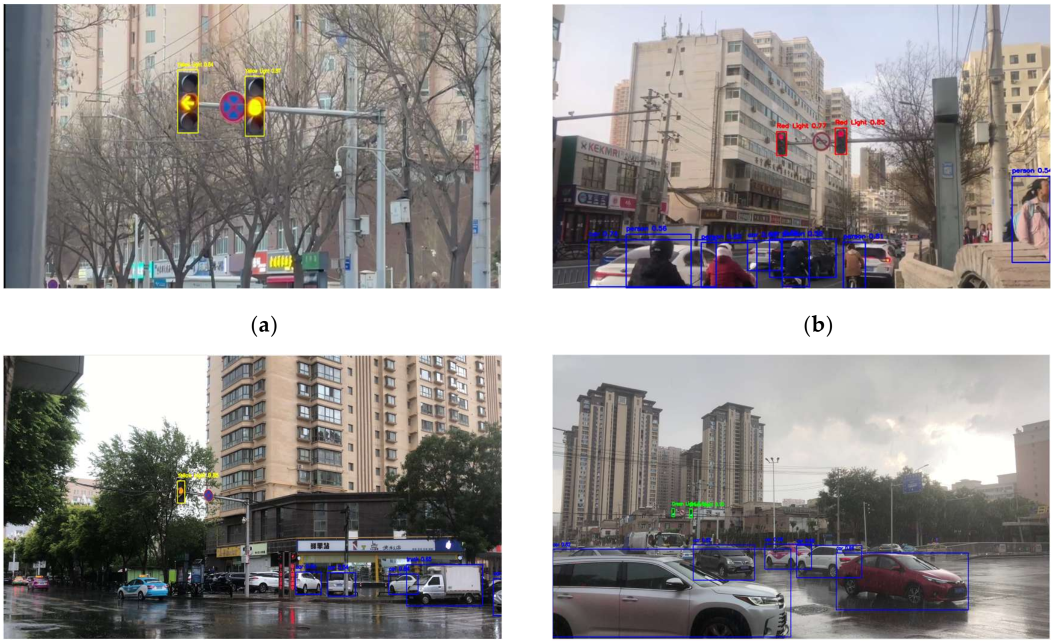
Figure 3. Comparison of YOLO 11n and YOLO 12n : (**a**) YOLO11 Confusion Matrix; (**b**) YOLO12 Confusion Matrix.

4. Results Analysis

This section presents evaluation of the proposed system, detection accuracy across diverse scenarios. The analysis is categorized into performance metrics and environmental robustness tests to verify the system's reliability for drivers with CVD.

4.1. Environmental Scenario Setup

Testing data are divided into daytime (Figure 4) and nighttime (Figure 5). Each time category was tested in different environmental conditions to determine the robustness and accuracy of the algorithm. In clear weather, natural lighting dominates, while in rainy weather, water droplets and reduced visibility interfere with detection performance. Both situations were further evaluated under normal and heavy traffic conditions, where vehicle density and obstacles can affect object visibility. A similar procedure was also applied to nighttime conditions. Nighttime testing was especially emphasized due to the difficulty of detecting traffic lights with dark or black frames in low-light conditions. In such scenarios, traffic light frame often blend into the background, leaving only the illuminated signals visible, which increases the risk of false detection or false negatives.



(**a**) (**b**)

(**c**) (**d**)

Figure 4. Daytime testing scenarios. (a) Clear weather, Normal traffic (DCN) ; (b) Clear weather, Crowded traffic (DCC); (c) Rainy weather, Normal traffic (DRN); and Rainy weather, Crowded traffic (DRC).

(**a**) (**b**)

(**c**) (**d**)

Figure 5. Nighttime testing scenarios. (a) Clear weather, Normal traffic (DCN) ; (b) Clear weather, Crowded traffic (DCC); (c) Rainy weather, Normal traffic (DRN); and Rainy weather, Crowded traffic (DRC).

4.2. Computational Efficiency and Real-Time Performance

The combination of various test scenarios was planned to evaluate system performance under diverse traffic conditions, ranging from ideal sunlight to challenging low-light and wet environments. A quantitative summary of average accuracy (confidence), processing speed (FPS), and inference time for each condition is shown in Table 2. Providing an objective basis for evaluating the computational efficiency of the proposed system.

Table 2. Summary of average system performance in various scenarios.

Scenario	FPS	Inference	Process
DCN	4.639	0.211	0.219
DCC	4.634	0.211	0.220
DRN	4.623	0.212	0.221
DRC	4.555	0.215	0.224
NCN	4.908	0.197	0.205
NCC	4.991	0.193	0.201
NRN	4.913	0.198	0.205
NRC	4.758	0.205	0.213

Average frames per second (FPS) of 4.75, average inference time of 0.205 seconds, and average processing time of 0.214 seconds. The amount of data acquired is based on the hardware used, since all training and inference tasks were done with a CPU.

4.3. Detection Accuracy and Model Robustness

Each of the eight environmental conditions is depicted in a 2–3 minute video (totaling 18 minutes) comprising footage sourced from various locations. The algorithm's detection performance for each dataset are shown in Table 3. The system was assessed under eight environmental conditions to thoroughly evaluate its performance based on these data.

Table 3. Detection performance metrics.

Metrics		**Daylight**				**Night**			
		Clear		**Rain**		**Clear**		**Rain**	
		Normal	**Crowded**	**Normal**	**Crowded**	**Normal**	**Crowded**	**Normal**	**Crowded**
Green Light Detected	Total	18672	22167	26059	29051	5456	9461	11818	14795
Red Light Detected	Total	10379	14166	15785	16372	2483	6235	6925	8066
Yellow Light Detected	Total	2306	2503	3542	3775	1160	1404	1692	2089
mAP Proxy	Avg	0.72	0.73	0.73	0.73	0.71	0.68	0.70	0.71
	Highest	0.95	0.95	0.95	0.95	0.9	0.90	0.9	0.91
Green Light Confidence	Avg	0.74	0.75	0.76	0.75	0.72	0.69	0.71	0.73
	Highest	0.95	0.95	0.95	0.95	0.91	0.91	0.91	0.92
Red Light Confidence	Avg	0.74	0.74	0.74	0.74	0.73	0.70	0.71	0.72
	Highest	0.9	0.9	0.9	0.9	0.89	0.89	0.89	0.89
Yellow Light Confidence	Avg	0.78	0.78	0.79	0.79	0.82	0.81	0.79	0.77
	Highest	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91

The performance of this system was evaluated based on its usage environment, namely on devices that only use CPUs. Due to variations in video duration, the number of recorded intersections, and traffic light cycle frequency, testing videos in different settings resulted in varying numbers of traffic light detections. For example, the Night–Clear–Normal (NCN) scenario produced the lowest number of detections at 9,099, while the Day–Rain–Crowded (DRC) scenario recorded the highest number at 49,198. Higher detection counts were naturally associated with longer recordings and more frequent traffic light cycles.

The reported mAP_proxy value in this analysis refers to a real-time proxy metric (hereafter denoted as mAP_proxy), defined as the average confidence score of detected traffic light instances during system operation. The maximum confidence level observed achived 0.95 across all detections. This metric does not correspond to the standard mAP@50 or mAP@50:95, as it does not rely on ground truth annotations or IoU-based precision–recall calculations. The highest confidence level were recorded during the day in clear weather and light traffic, while the lowest values were recorded in more complex scenarios, such as Night–Rain–Crowded (NRC).

4.3. Spatial Robustness and Signal Configuration Analysis

In addition to standard circular traffic lights, this system is evaluated based on its ability to recognize traffic lights with different orientations and shapes, such as horizontal configurations and directional signals. The system correctly interprets the active signal based on its position shown in Figure 6.

(**a**) (**b**) (**c**)

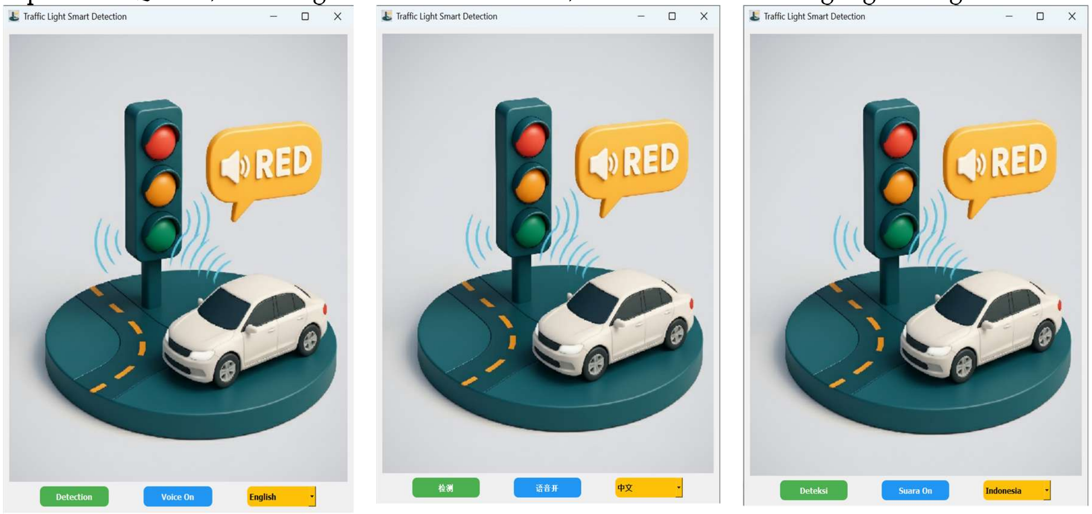
Figure 6. Detection performance across diverse configurations: (a) Vertical frame detection; (b) Horizontal frame detection; (c) Recognition of active status in arrow-shaped signals.

Critically, this system does not rely entirely on color information. It utilizes the spatial and structural configuration of traffic lights. During the detection process, the model recognizes the entire traffic light array, and the active signal is determined based on the relative position of the lit lights in the frame (top–middle–bottom for vertically arranged lights and left–middle–right for horizontally arranged lights).

The advantage of this system lies in its spatial positioning logic, rather than simply shape classification. By detecting the entire traffic light frame, this model identifies the active status of the signal based on its position relative to the frame. In horizontal configurations, the system correctly interprets active signals based on their position. The system points out exactly which part of the traffic light is active, effectively translating the arrow's meaning through its position in the frame.

5. System Implementation and Functional Discussion

The development of the Traffic Light Smart Detection application adheres to the procedure depicted in Figure 7. The system incorporates detection through YOLO models, interactive visual output via a QLabel, multilingual audio instructions, and user interface language configurations.



(**a**) (**b**) (**c**)

Figure 7. Initial Interface of the Traffic Light Smart Detection application. The application is presented in three language options: (**a**) English, (**b**) Mandarin, and (**c**) Indonesian.

When users turn on detection, the app uses OpenCV to open a video stream and starts a detection loop with QTimer. This makes sure that the frame update process happens on a regular basis without slowing down the interface, as illustrated in Figure 8. After being preprocessed, each frame is given to two YOLO models. YOLOv12 traffic light custom model looks for the status of traffic lights, and the original of YOLOv12 looks for other things that are important for analyzing traffic density. The outcomes of the detection include bounding boxes, class labels, and model confidence. The bounding box visualization updates and if the sound feature is turned on, the system makes a sound warning based on the color of the light that is detected. The sound control system also has a cooldown feature that stops the system from giving voice instructions too many times. The detection goes on until the user presses the StopDetection button, which pauses the QTimer and frees up the camera source.

Figure 8. Application Runtime Detection.

4.1. *Visual Perception*

The camera installed in the car will capture video in real time when the detection function is activated. The input data is then forwarded to the system to be processed frame by frame. Only predictions with a confidence level of more than 0.50 are retained so that low detections are not displayed. Detected objects are displayed in dark blue boxes, while traffic light objects are displayed according to the detected color. Level of detection confidence and object name in the form of a text label are included to provide additional information. The model uses a timer mechanism to perform inference on each frame. This process is repeated periodically to ensure that no images are missed. Detection updates and frame display run synchronously and stably so that users see results that are consistent with what the system processes. The visual design is minimalist to avoid information overload and to make it easier for users to use as a driving aid.

4.2. *Audio Feedback*

The auditory feedback instructions will be delivered to the user based on detection results. This implementation uses a Python library that enables local text-to-speech (TTS), allowing audio to be produced directly through the system speaker without requiring an internet connection or external audio files. Warning messages for each traffic light condition are generated dynamically through the local TTS module (pyttsx3), using the computer's operating system voice. The system generates distinct audio instructions corresponding to each light status as follows:

- Red Light: “Red light, please stop and relax!”
- Yellow Light: “Yellow light, please prepare!”
- Green Light: “Green light, you may go, have a pleasent journey!”

To reduce continuous repetition of sounds when frames are detected, audio is only activated when there is a change in traffic light signal detection, or from undetected to detected. As long as the light status does not change even though the object remains detected in many consecutive frames, the sound is not played repeatedly. A minimum time limit of 1 second is also applied between sound playbacks. Even if there is a very rapid change in status, the system will wait for this time interval before allowing the next sound to play. Reducing a user’s cognitive fatigue from becoming mentally tired due to excessive or repetitive information by overlapping sounds. Object detection continues to be displayed in real time, without slowing down due to sound output. This is because sound is processed in a separate threads so that detection remains smooth.

4.3. Multilingual

The system supports multilingualism, which is proposed to help users from different backgrounds use the application. Language options are available in English, Mandarin, and Indonesian, including text and audio feedback. Before starting detection or during the detection process, users are free to choose the language according to their needs without having to turn off the application. All parts of the system can adjust the interface text, brief explanations or tooltips, and audio instructions. Users will not see a mixture of languages because the system works consistently. Each language has its own text and audio pair. The selected language does not only affect the text but also the voice used.

6. Conclusion

6.1. Main Contributions

This study offers a solution for drivers with color vision deficiency assistance systems with voice feedback in recognizing traffic lights. The experimental analysis with the YOLO custom model traffic light detection obtained the following result:

- After comparing with previous model training results, this study decided to use YOLOv12 for implementation in the user application;
- The model was tested with eight scenarios involving heavy conditions at night, bad weather, and crowded traffic. During the day, the highest confidence level achieved 0.95, while at night the detection performance decreases, especially in crowded environments;
- The application is designed with traffic light visualization that included bounding boxes, name labels, and confidence levels. The audio produced followed the user’s language selection and are only played when there is a change in traffic light status.

6.1. System Limitations and Safety Considerations

Methods that rely entirely on color are prone to classification errors if the system is trained only through annotations on active lights without labeling the entire traffic light frame. By prioritizing the pos ition and structure of active lights, the proposed system can reduce the risk of misinterpretation, including critical false negatives such as recognizing car backlights or street lights as traffic lights in nighttime conditions with heavy traffic density. The main problem is that the effectiveness of full detection depends on the vehicle’s camera. If the image input is poor or the resolution is low, the system will have difficulty capturing the full details of the traffic light structure. As a result, the risk of detection failure becomes higher.

In some conditions where red and yellow lights are lit simultaneously in a same frame of traffic light, the system has difficulty determining signal priority, causing detection ambiguity. We still need to refine the algorithm so that it does not get confused when faced with such double signals. In more complex scenarios where false negatives may occur, such as Night–Rain–Crowded conditions, the absence of system output should be interpreted as a cue for increased driver vigilance rather than as an indication to proceed, reinforcing the role of the system as a supplementary assistance tool.

6.3. Future Work

Based on the results of this prototype, future development can focus on the model training process by analyzing the confusion matrix to evaluate detection errors during model training. Optimize the algorithm to improve detection accuracy and speed. In addition, expanding the variety of datasets related to distance, position, size, and perspective of traffic lights is also very important. Further system development could combine the model with additional sensors, especially when dealing with extreme weather conditions and complex environments.

Direct integration of the system into vehicles, including comparisons of the use of cameras inside or outside the car. The system also needs to support more flexible natural language settings, with additional language options for users from different backgrounds. A user-friendly interface design for individuals with color vision deficiency needs to be designed for ease of use. Finally, hands-on testing for users with color vision deficiency is necessary to ensure the effectiveness of the system.

Author Contributions: Conceptualization, M.Y.; Methodology, F.A., Q.A., and Z.K.; Software, F.A.; Investigation, Q.A.; Data curation, F.A. and Z.K.; Formal analysis, M.Y., F.A., and L.B.; Validation, M.Y., F.A., Q.A., L.B., Z.K., and M.T.; Visualization, L.B.; Resources, F.A., Q.A., and L.B.; Writing—original draft preparation, M.Y. and F.A.; Writing—review and editing, M.Y., Q.A., Z.K., and M.T.; Supervision, M.Y. and M.T.; Project administration, M.Y.; Funding acquisition, M.Y. and M.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (NSFC), grant number 62363022.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets generated during this study are publicly available in the Zenodo repository at https://zenodo.org/records/18039880.

Acknowledgments: The authors would like to acknowledge the use of Label Studio to annotate datasets, generative AI tools, including GPT, Gemini, Runway, QuillBot, DeepL, and Canva, for assistance in language refinement, content editing, and visualization during the preparation of this manuscript. The authors have reviewed and edited all generated outputs and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

Avg	Average
CVD	Color Vision Deficiency
CNN	Convolutional Neural Network
DL	Deep Learning
FPS	Frames Per Second
GPU	Graphics Processing Unit
HSV	Hue, Saturation, Value
i7	Intel Core i7 Processor

mAP	Mean Average Precision
mAP proxy	The average confidence score of detected traffic light
NCN	Night, Clear,Normal traffic scenario
NCC	Night, Clear, Crowded traffic scenario
NRN	Night, Rain, Normal traffic scenario
NRC	Night, Rain, Crowded traffic scenario
DCN	Day, Clear, Normal traffic scenario
DCC	Day, Clear, Crowded traffic scenario
DRN	Day, Rain, Normal traffic scenario
DRC	Day, Rain, Crowded traffic scenario
QTimer	Qt Timer (PyQt5)
SGD	Stochastic Gradient Descent
TTS	Text-to-Speech
YOLO	You Only Look Once

References

1. Simunovic, M.P. Colour Vision Deficiency. *Eye* **2010**, *24*, 747–755, doi:10.1038/eye.2009.251.
2. Almustanyir, A. A Global Perspective of Color Vision Deficiency: Awareness, Diagnosis, and Lived Experiences. *Healthcare* **2025**, *13*, 2031, doi:10.3390/healthcare13162031.
3. Khatri, A.; K.C, B.K.; Gautam, S.; Kharel, M. The Burden of Color Vision Defect in Nepal – an Iceberg Phenomenon. *J Chitwan Med Coll* **2019**, *9*, 69–71, doi:10.3126/jcmc.v9i4.26906.
4. Nazeer, M.; Bashir, S.; Rafiq, N. Color Vision Deficiency in Medical Students in Jammu & Kashmir, India. *Galician med. j.* **2019**, *26*, doi:10.21802/gmj.2019.1.9.
5. Tagarelli, A.; Piro, A.; Tagarelli, G.; Lantieri, P.B.; Risso, D.; Olivieri, R.L. Colour Blindness in Everyday Life and Car Driving. *Acta Ophthalmologica Scandinavica* **2004**, *82*, 436–442, doi:10.1111/j.1395-3907.2004.00283.x.
6. Kim, Y.K.; Kim, K.W.; Yang, X. Real Time Traffic Light Recognition System for Color Vision Deficiencies. In Proceedings of the 2007 International Conference on Mechatronics and Automation; IEEE: Harbin, China, August 2007; pp. 76–81.
7. Pepple, G.; Adio, A. Visual Function of Drivers and Its Relationship to Road Traffic Accidents in Urban Africa. *SpringerPlus* **2014**, *3*, 47, doi:10.1186/2193-1801-3-47.
8. Nuñez, J.R.; Anderton, C.R.; Renslow, R.S. Optimizing Colormaps with Consideration for Color Vision Deficiency to Enable Accurate Interpretation of Scientific Data. *PLoS ONE* **2018**, *13*, e0199239, doi:10.1371/journal.pone.0199239.
9. Rocchini, D.; Nowosad, J.; D'Introno, R.; Chieffallo, L.; Bacaro, G.; Gatti, R.C.; Foody, G.M.; Furrer, R.; Gábor, L.; Malavasi, M.; et al. Scientific Maps Should Reach Everyone: The Cblindplot R Package to Let Colour Blind People Visualise Spatial Patterns. *Ecological Informatics* **2023**, *76*, 102045, doi:10.1016/j.ecoinf.2023.102045.
10. Tan, T.F.; Wongsawad, W.; Hurairah, H.; Loy, M.J.; Lwin, W.W.; Mohd Rawi, N.A.; Sidik, M.; Grzybowski, A.; Raman, R.; Ruamviboonsuk, P.; et al. Colour Vision Restrictions for Driving: An Evidence-Based Perspective on Regulations in ASEAN Countries Compared to Other Countries. *The Lancet Regional Health - Southeast Asia* **2023**, *14*, 100171, doi:10.1016/j.lansea.2023.100171.
11. Nasruddin, N.I.; Arimaswati, A.; Dewi Nughrawati Putri; Rustam Hn, M. DETEKSI BUTA WARNA DENGAN METODE ISHIHARA PADA MAHASISWA BARU JALUR PENERIMAAN SNMPTN UNIVERSITAS HALU OLEO. *J-ABDI* **2023**, *2*, 6923–6928, doi:10.53625/jabdi.v2i11.5377.
12. Kobal, N.; Hawlina, M. Comparison of Visual Requirements and Regulations for Obtaining a Driving License in Different European Countries and Some Open Questions on Their Adequacy. *Front. Hum. Neurosci.* **2022**, *16*, 927712, doi:10.3389/fnhum.2022.927712.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 January 2026 doi:10.20944/preprints202601.0608.v1

13. GOV.UK. Visual Disorders: Assessing Fitness to Drive. **2025**. Available online: https://www.gov.uk/guidance/visual-disorders-assessing-fitness-to-drive#colour-blindness (accessed on 31 December 2025).
14. Lin, H.-Y.; Chen, L.-Q.; Wang, M.-L. Improving Discrimination in Color Vision Deficiency by Image Re-Coloring. *Sensors* **2019**, *19*, 2250, doi:10.3390/s19102250.
15. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444, doi:10.1038/nature14539.
16. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); IEEE: Las Vegas, NV, USA, June 2016; pp. 779–788.
17. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; IEEE: Salt Lake City, UT, June 2018; pp. 4510–4520.
18. Behrendt, K.; Novak, L.; Botros, R. A Deep Learning Approach to Traffic Lights: Detection, Tracking, and Classification. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA); IEEE: Singapore, Singapore, May 2017; pp. 1370–1377.
19. Pavlitska, S.; Lambing, N.; Bangaru, A.K.; Zöllner, J.M. Traffic Light Recognition Using Convolutional Neural Networks: A Survey 2023.
20. Che, M.; Che, M.; Chao, Z.; Cao, X. Traffic Light Recognition for Real Scenes Based on Image Processing and Deep Learning. *cai* **2020**, *39*, 439–463, doi:10.31577/cai_2020_3_439.
21. Zeng, Y.; Lan, J.; Ran, B.; Wang, Q.; Gao, J. Restoration of Motion-Blurred Image Based on Border Deformation Detection: A Traffic Sign Restoration Model. *PLoS ONE* **2015**, *10*, e0120885, doi:10.1371/journal.pone.0120885.
22. Possatti, L.C.; Guidolini, R.; Cardoso, V.B.; Berriel, R.F.; Paixao, T.M.; Badue, C.; De Souza, A.F.; Oliveira-Santos, T. Traffic Light Recognition Using Deep Learning and Prior Maps for Autonomous Cars. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN); IEEE: Budapest, Hungary, July 2019; pp. 1–8.
23. Nine, J.; Mathavan, R. Traffic Light and Back-Light Recognition Using Deep Learning and Image Processing with Raspberry Pi. *Embedded Selforganising Systems* **2021**, *8*, 1519, doi:10.14464/ESS.V8I2.490.
24. Manawadu, M.; Wijenayake, U. Voice-Assisted Real-Time Traffic Sign Recognition System Using Convolutional Neural Network 2024.
25. Sukhani, K.; Shankarmani, R.; Shah, J.; Shah, K. Traffic Sign Board Recognition and Voice Alert System Using Convolutional Neural Network. In Proceedings of the 2021 2nd International Conference for Emerging Technology (INCET); IEEE: Belagavi, India, May 21 2021; pp. 1–5.
26. Dewangan, R.K.; Chaubey, D.S. Object Detection System with Voice Output Using Python. **2021**, *6*.
27. Kompalli, P.L.; Kalidindi, A.; Chilukala, J.; Nerella, K.; Shaik, W.; Cherukuri, D. A Color Guide for Color Blind People Using Image Processing and OpenCV. *Int. J. Onl. Eng.* **2023**, *19*, 30–46, doi:10.3991/ijoe.v19i09.39177.
28. Goenawan, A.D.; Rachman, M.B.A.; Pulungan, M.P. Identifikasi Warna Pada Objek Citra Digital Secara Real Time Menggunakan Pengolahan Model Warna HSV. *JURTIE* **2022**, *4*, 68–74, doi:10.55542/jurtie.v4i1.430.
29. Vel Tech High Tech Dr.Rangarajan Dr.Sakunthala Engineering College, Avadi; Nevathetha, R.A.; Fathima, K.K.; Niveditha, S.; Selvavathi, M. COLOR DETECTION USING OPENCV. *IJSREM* **2023**, *07*, doi:10.55041/IJSREM17740.
30. Zheng, Z.; Cheng, J.; Jin, F. TFP-YOLO: Obstacle and Traffic Sign Detection for Assisting Visually Impaired Pedestrians. *Sensors* **2025**, *25*, 5879, doi:10.3390/s25185879.
31. Nguyen, T. Improve Underwater Object Detection through YOLOv12 Architecture and Physics-Informed Augmentation 2025.
32. Tian, Y.; Ye, Q.; Doermann, D. YOLOv12: Attention-Centric Real-Time Object Detectors 2025.
33. Shivakumar, N. Colored Object Detection For Blind People Using CNN. **2022**, *10*.
34. Yagob, F.; Sasiadek, J.Z. Enhanced Real-Time Method Traffic Light Signal Color Recognition Using Advanced Convolutional Neural Network Techniques. *WEVJ* **2025**, *16*, 441, doi:10.3390/wevj16080441.

35. Khaled, L.B.; Rahman, M.; Ebu, I.A.; Ball, J.E. FlashLightNet: An End-to-End Deep Learning Framework for Real-Time Detection and Classification of Static and Flashing Traffic Light States. *Sensors* **2025**, *25*, 6423, doi:10.3390/s25206423.
36. Gong, C.; Li, A.; Song, Y.; Xu, N.; He, W. Traffic Sign Recognition Based on the YOLOv3 Algorithm. *Sensors* **2022**, *22*, 9345, doi:10.3390/s22239345.
37. Wang, Q.; Zhang, Q.; Liang, X.; Wang, Y.; Zhou, C.; Mikulovich, V.I. Traffic Lights Detection and Recognition Method Based on the Improved YOLOv4 Algorithm. *Sensors* **2021**, *22*, 200, doi:10.3390/s22010200.
38. Hindarto, D. Enhancing Road Safety with Convolutional Neural Network Traffic Sign Classification. *SinkrOn* **2023**, *8*, 2810–2818, doi:10.33395/sinkron.v8i4.13124.
39. Najm, H.; Elferjani, K.; Alariyibi, A. Assisting Blind People Using Object Detection with Vocal Feedback. In Proceedings of the 2022 IEEE 2nd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA); IEEE: Sabratha, Libya, May 23 2022; pp. 48–52.
40. Das, D.; Roy, S. Object Detection with Voice Output for Visually Impaired. In Proceedings of the 2024 International Conference on Communication, Computing and Internet of Things (IC3IoT); IEEE: Chennai, India, April 17 2024; pp. 1–6.
41. Ravindra Karmarkar, R.; Honmane, Prof.V.N. OBJECT DETECTION SYSTEM FOR THE BLIND WITH VOICEGUIDANCE. *IJEAST* **2021**, *6*, doi:10.33564/IJEAST.2021.v06i02.013.