Detection and Classification of River and Lake Signs Using the YOLO Algorithm

Raden Muhamad Firzatullah[1*], Moh. Aziz Rohman[1], Donny Afrizal Melayu[1], Arleiny[2], Rizal Eko, Andhika Oktariansyah[1]

[1] Politeknik Transportasi Sungai, Danau dan Penyeberangan Palembang, Indonesia

[2]Politeknik Pelayaran Surabaya, Indonesia

Email: firza.bogor@gmail.com

Abstract : Waterway signs in rivers and lakes play a vital role in ensuring navigation safety by providing directional guidance, hazard warnings, speed limitations, and safe passage indicators for vessels. Manual inspection and monitoring systems currently used in Indonesia are time-consuming, expensive, and vulnerable to delays in detecting damaged or missing signs due to limited inspection frequency and environmental constraints. Meanwhile, the dynamic and complex characteristics of inland waterways — including small sign size in images, varying distances, fluctuating lighting conditions, and cluttered backgrounds — pose significant challenges for automated detection systems. To address these issues, this study proposes the utilization of the YOLO (You Only Look Once) algorithm for automatic detection and classification of river and lake navigation signs. YOLO is known for its speed and accuracy in real-time object detection and has demonstrated strong performance in various domains, including maritime applications. Despite its potential, research on the automated detection of inland waterway signs, particularly in Indonesia, remains limited. This study aims to develop an efficient and robust computer-vision-based system capable of identifying inland waterway signs in accordance with national regulations. The expected outcomes include improved monitoring efficiency, enhanced navigation safety through early detection of missing or damaged signs, and the creation of a localized image dataset to support future research and the advancement of artificial intelligence in aquatic transportation safety.

Keywords: YOLO, inland waterway signs, computer vision

1. INTRODUCTION

Waterway signs, both in rivers and lakes, are essential elements for ensuring navigational safety. These signs provide information on sailing directions, hazard warnings, speed limitations, and safe passage routes for vessels (Ministry of Transportation, 2011; Ministry of Transportation, 2012). Missing, damaged, or displaced signs can lead to accidents such as collisions or groundings. Therefore, monitoring and inspecting the condition of waterway signs are vital aspects of navigation safety management in Indonesia's inland waterways.

A. Yulianto et al. (eds.), *Proceedings of the International Conference of Inland Water and Ferries Transport Polytechnic of Palembang on Technology and Environment (IWPOSPA-TE 2025)*, Advances in Engineering Research 302,
https://doi.org/10.2991/978-94-6239-731-6_8

To date, the collection and monitoring of waterway sign data have been conducted manually by field officers through direct inspection or periodic reporting. This method is time-consuming, costly, and dependent on weather conditions and site accessibility. Moreover, infrequent inspections often result in delayed detection of damaged or missing signs (Flores-Calero et al., 2024). With the increasing need for digitalization and automation in water transportation, computer-vision-based approaches offer a promising solution for automatically detecting and classifying signs from images or videos.

However, detecting waterway signs presents unique challenges compared to road traffic signs. The small size of signs in images, varying viewing distances, changing lighting conditions due to water reflections, and complex backgrounds such as vegetation and other vessels can reduce detection accuracy (McMillan et al., 2023). Additionally, weather conditions such as fog and rain can degrade image quality. Therefore, a detection method is required that not only operates quickly but is also robust to variations in waterway environments.

One of the most widely used and proven effective algorithms for real-time object detection is YOLO (You Only Look Once). First introduced by Redmon et al. (2016), YOLO has continued to evolve into its latest variants such as YOLOv5 and YOLOv8 (Ultralytics, 2024). YOLO offers advantages in speed and efficiency by performing detection and classification in a single stage, unlike two-stage methods such as Faster R-CNN. This approach enables real-time system operation even on low-power computing devices, such as edge devices (e.g., NVIDIA Jetson or Raspberry Pi).

Previous studies have shown that the YOLO algorithm performs well in detecting signs and visual markers across various domains. Flores-Calero et al. (2024), in their systematic review, reported that YOLO consistently demonstrates high accuracy and fast inference times for traffic sign detection. Another study by McMillan et al. (2023) demonstrated the successful application of YOLO with transfer learning for buoy detection in shellfish aquaculture automation systems, achieving significant accuracy improvements compared to baseline models. These findings indicate that YOLO can be effectively adapted to maritime and aquatic contexts.

Nevertheless, research specifically addressing the detection of inland waterway signs (rivers and lakes) remains very limited, particularly in Indonesia. Inland waterway signs differ from marine and road traffic signs in terms of shape, color, and installation placement. This highlights a research gap in developing automated computer-vision systems capable of recognizing and classifying waterway signs according to national regulations (Ministry of Transportation, 2011).

Therefore, this study is crucial to develop a YOLO-based detection and classification system for river and lake navigation signs. The system is expected to support automatic, efficient, and accurate monitoring of waterway signs, contributing to navigation safety digitalization within the Ministry of Transportation. Furthermore, this research will produce a localized image dataset of waterway signs to support future studies and serve as an initial step toward adopting Artificial Intelligence (AI) in water transportation safety.

2. METHODOLOGY

2.1. RESEARCH FRAMEWORK

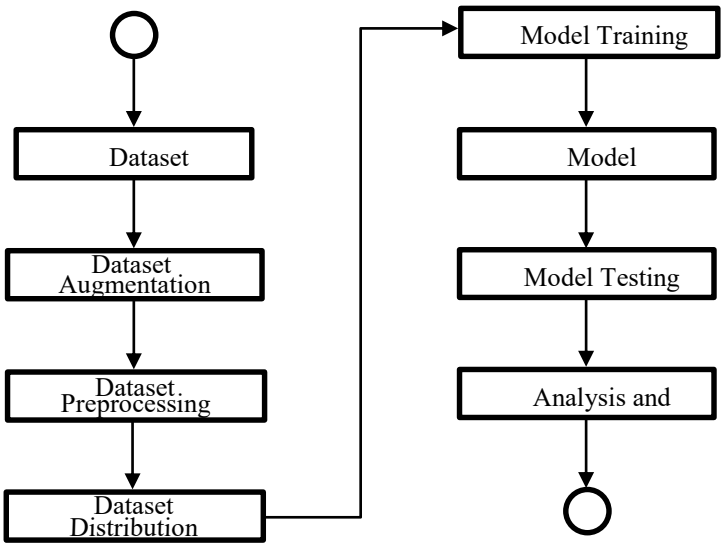

Figure 1. Research Framework

This research consists of several main stages, beginning with dataset collection, which involves obtaining images of river and lake signs through field photography and relevant secondary sources. The next stage is dataset augmentation to increase data variation using techniques such as rotation, flipping, and lighting adjustments to enhance the model's generalization capability. After that, dataset preprocessing is conducted, including image resizing, pixel normalization, and converting labels into the YOLO format to prepare the data for training.

The dataset is then divided during the dataset distribution phase into three parts: training, validation, and testing, using a specified proportion. The next step is model training using the YOLOv7 algorithm to learn visual patterns of waterway signs, followed by model validation to monitor performance and adjust parameters to prevent overfitting. Once the optimal model is obtained, model testing is carried out using new data to evaluate detection performance based on metrics such as precision, recall, and mean Average Precision (mAP). The final stage, analysis and evaluation, aims to assess the testing results, identify model limitations, and draw conclusions regarding the effectiveness of YOLOv7 in detecting and classifying river and lake navigation signs.

This technique for presenting processed data uses an interactive model which includes three analysis components, namely reduction, data presentation and drawing conclusions (Miles and Huberman, 1992:20).

2.2. DATA COLLECTING

This study uses 150 original images of river signs consisting of six label classes referring to the provisions of the Ministry of Transportation Regulation Number 52 of 2012 concerning River and Lake Navigation Channels. Each label represents a different

sign type corresponding to prohibition, mandatory, and warning categories. To increase variation and improve the model's robustness to various environmental conditions, data augmentation was performed by adding effects such as low resolution, noise, blur, lighting artifacts, occlusion, and cluttered background conditions. This augmentation aims to enable the YOLOv7 model to adapt to images with varying quality and lighting conditions similar to those encountered in real field environments. After the augmentation process, the total dataset used for training, validation, and testing increased to 1,050 images, which is expected to enhance the model's generalization ability in accurately detecting and classifying river navigation signs.

2.3. Dataset Augmentation

In an effort to increase training data under various conditions of visual degradation, each image in the dataset was transformed to represent situations such as low resolution, noise, blur, lighting disturbances, occlusion, and complex backgrounds (cluttered backgrounds). Low-resolution images were generated using a downsampling method (Bruckstein et al., 2023), while noise was added using the salt-and-pepper method (Kaur, 2014). Blurred conditions were simulated by applying Gaussian Blur (Singh & Kaur, 2015), whereas lighting intensity variations were adjusted using Gamma Correction (Huang & Zhang, 2012). To mimic partial occlusion, the Random Erasing method was employed (Zhang et al., 2020), and to create busy or irregular background conditions, the CutMix technique was applied (Yun et al., 2019).

2.4. Dataset Preprocessing

In the preprocessing stage, all images in the dataset were resized to a fixed resolution of 512×512 pixels to meet the requirements of object detection architectures such as SSD and YOLOv7, which demand a constant input size to ensure consistent feature extraction and computational efficiency (Redmon et al., 2016; Liu et al., 2016). After standardizing the image dimensions, object annotation was performed, involving marking the position of objects with bounding boxes and assigning labels according to the predefined categories. This step plays a crucial role as supervised learning guidance for the model during the training phase, enabling the system to accurately recognize and classify objects (Everingham et al., 2010).

2.5. Dataset Distribution

In this study, the dataset was divided using the K-Fold Cross Validation method, an evaluation approach that splits the dataset into balanced folds to ensure comprehensive testing and prevent overfitting. A total of 1,050 images in the dataset were separated into three main groups—training data, validation data, and testing data—with proportions of 70%, 10%, and 20%, respectively, as shown in Table 1. This method was chosen because it optimizes the use of all available data, produces consistent model evaluation results, and enhances the generalization capability of the developed model (Kohavi, 1995; Refaeilzadeh, Tang, & Liu, 2009).

2.6. MODELING

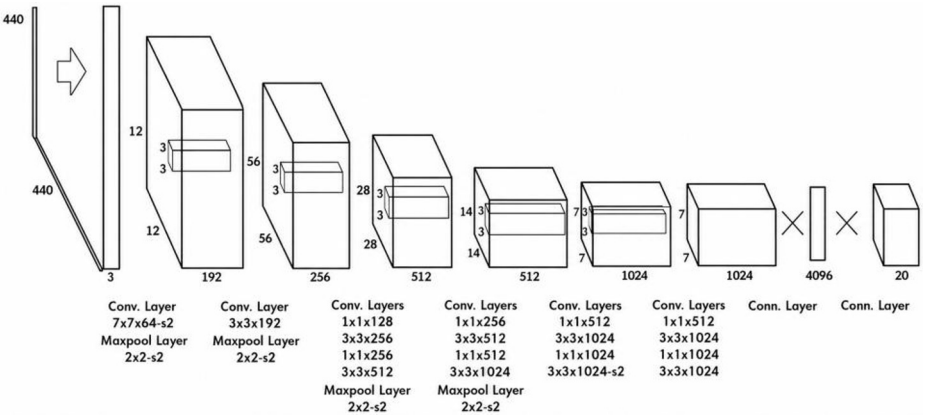

Figure 2. YOLOv7 Architecture (Shehata and El-Sayed, 2018)

The YOLO (You Only Look Once) architecture, as illustrated in the figure, is a convolutional neural network structure designed to perform end-to-end object detection in a single processing stage. The model receives an input image with a resolution of 448×448 pixels and three color channels (RGB). The initial process begins with a 7×7 convolutional layer with 64 filters and a stride of 2, followed by a 2×2 max-pooling layer that extracts basic features from the image.

Next, the network consists of several 3×3 convolutional layers with an increasing number of filters ranging from 192 to 1024, combined with max-pooling layers to gradually reduce spatial dimensions while preserving important visual information. Deeper convolutional layers generate more complex feature representations, enabling the network to recognize objects of various sizes and shapes.

After the feature extraction stage, the output is passed into two fully connected layers, one of which contains 4096 nodes, to perform classification and bounding-box regression. The final layer produces a prediction vector—30 in this case—containing information about object positions, detection confidence scores, and object classes.

This architecture was used in the study by Shehata and El-Sayed (2018) to detect and localize human actions, where YOLO was chosen due to its ability to perform simultaneous and efficient detection suitable for real-time applications. The one-stage approach in YOLO enables both object classification and localization to be performed directly within a single integrated network, unlike two-stage methods such as R-CNN, which separate these processes.

2.7. ANALYSIS AND EVALUATION

The performance evaluation of a classification system typically uses a confusion matrix, which is a 2 × 2 matrix that describes the model's prediction outcomes on test data in binary classification cases (Dewi et al., 2024). A brief explanation of the components in the confusion matrix is presented in Table 2 to provide a basic understanding of how to interpret classification results.

Tabel 2. Confusion Matrix

		Prediction Category	
		Positive	**Negative**
Actual Category	**Positive**	TP	FN
	Negative	FP	TN

As shown in Table 1, the True Positive (TP) component represents data that is predicted correctly as positive and is indeed positive in reality, whereas True Negative (TN) represents data that is predicted as negative and is actually negative. False Positive (FP) describes cases in which the model predicts positive for data that is actually negative, while False Negative (FN) indicates negative predictions for data that is actually positive. The values of precision and recall are derived from the relationship between TP, FP, and FN. Meanwhile, accuracy is used to assess the overall performance of the model based on the proportion of correct predictions to the total number of predictions made, and it is calculated using Equation (4).

$$Akurasi = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Presisi = \frac{TP}{TP + FP}$$

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

More specifically, precision measures how many of the predicted positive results are truly relevant out of all data predicted as positive (Andika, 2019), while recall indicates the proportion of relevant data correctly identified by the model out of all actual relevant data (Prasetyawan et al., 2018). The calculation is performed based on classification results on the test data, and to determine valid or invalid categories, repeated predictions are conducted so that the results can be presented in the form of graphical diagrams that visually illustrate the model's performance.

3. Result and Discussion

3.1. Model Accuration

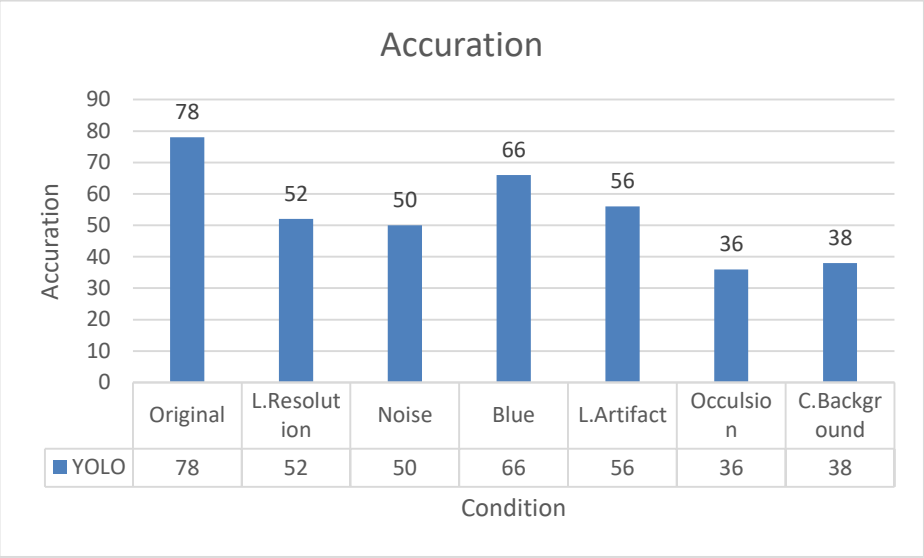

Figure 3. Model Accuration

Figure 3 presents the test results shown in the previous figure, where the YOLOv7 model demonstrates varying accuracy levels in detecting river signs under different image conditions. In the original (normal) image condition, the model achieved the highest accuracy of 78%, indicating that YOLOv7 can detect objects effectively when image quality is undisturbed. However, when the image quality decreases, such as in low-resolution (52%) and noisy (50%) images, the detection accuracy drops significantly. This occurs because disturbances in the image reduce the clarity of visual features used by the model for classification.

Additionally, in blurred image conditions, the model achieved 66% accuracy, indicating that despite reduced sharpness, YOLOv7 is still capable of recognizing the shape patterns of signs fairly well. Meanwhile, in low-artifact conditions, which introduce image distortion, an accuracy of 56% reflects the model's sensitivity to changes in object shapes or contours. Occlusion conditions produced the lowest accuracy of 36%, followed by complex backgrounds at 38%, indicating that the model struggles to recognize objects when parts of the sign are covered or the background is visually complex.

Overall, these results indicate that image quality has a significant effect on YOLOv7 detection performance. The model performs optimally on clear images with minimal disturbance, while poor lighting conditions, low resolution, and occluded objects cause a substantial performance decline. Therefore, it can be concluded that improving image acquisition quality and applying image preprocessing techniques are important factors in increasing the accuracy of YOLOv7-based river sign detection systems.

3.2. Model Precision

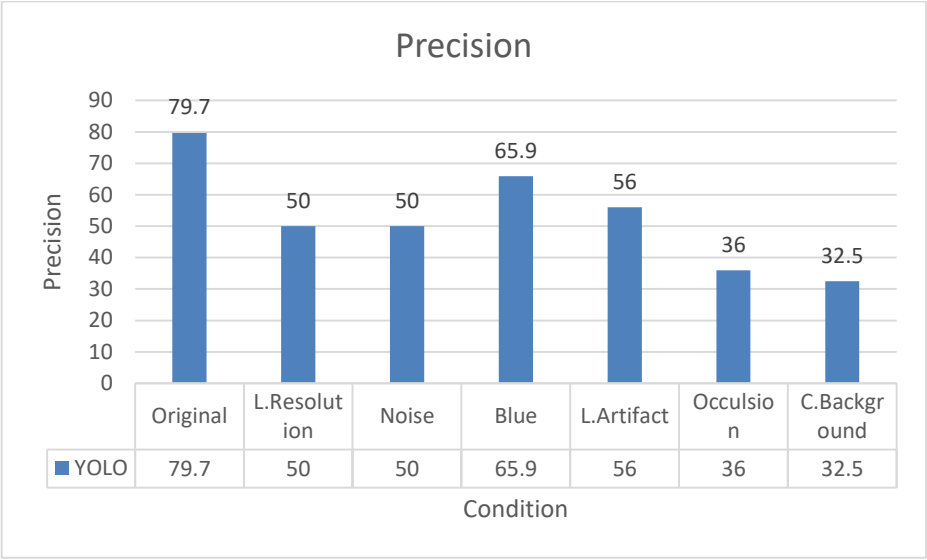

Figure 4. Model Precision

Based on the test results shown in the figure above, the precision value of the YOLOv7 model in detecting river signs exhibits considerable variation across different image conditions. In the original (normal) condition, the model achieves the highest precision of 79.7%, indicating its ability to minimize false positive detections when the image is in ideal condition. However, in low-resolution and noisy image conditions, the precision value drops drastically to 50%, indicating that visual disturbances such as reduced pixel quality and noise hinder the model's ability to recognize key features of river sign objects.

In blurred image conditions, the precision value remains fairly good at 65.9%, showing that the model can still detect objects even when slight blurring occurs. The low-artifact condition results in a precision value of 56%, suggesting that image distortion still affects the model's prediction accuracy. The most significant decrease is observed under occlusion and complex background conditions, with precision values of 36% and 32.5%, respectively, indicating that when parts of the object are covered or the background is complex, the model's ability to accurately detect objects becomes highly limited.

Overall, these findings demonstrate that image quality and object clarity greatly influence the precision value of YOLOv7. The model performs optimally on images with good lighting and high clarity but experiences a significant performance decline on visually degraded images. These results highlight the importance of image preprocessing steps—such as contrast enhancement, noise reduction, or lighting adjustment—to improve the reliability of river-sign detection systems based on YOLOv7.

3.3. Model Recall

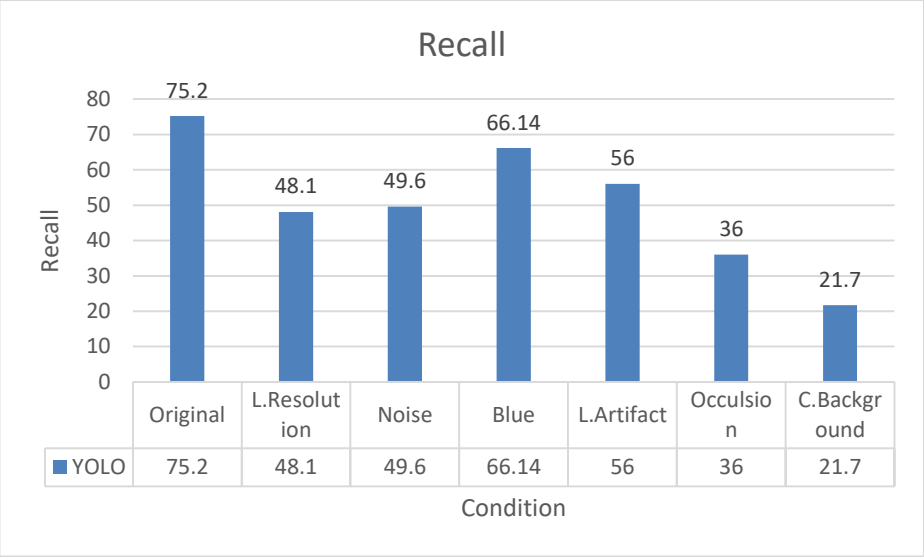

Figure 5. Model Recall

Based on the test results shown in the figure above, the recall value of the YOLOv7 model in detecting river signs demonstrates considerable variation across different image conditions. In the original (normal) condition, the model achieves the highest recall value of 75.2%, indicating its capability to correctly identify most river sign objects present in clean images without visual disturbances. However, under low-resolution and noisy conditions, the recall value decreases to 48.1% and 49.6% respectively, suggesting that reduced image quality or pixel interference causes the model to miss some objects that should have been detected.

Under blurred conditions, the recall reaches 66.14%, showing that the model remains fairly capable of recognizing relevant objects despite image blur. The low-artifact condition yields a recall value of 56%, indicating that image distortion still affects detection performance, though not as severely. Meanwhile, the occlusion and complex-background conditions show the lowest recall values, at 36% and 21.7% respectively, meaning the model struggles to recognize objects when parts of the sign are covered or when the object is mixed with a complex background.

Overall, these results indicate that the recall performance of YOLOv7 is highly dependent on object clarity and visibility within the image. The model performs optimally on images with good lighting and without disturbances but experiences significant performance degradation when objects are occluded, distorted, or blended with complex environments. This highlights the importance of applying image enhancement techniques and incorporating diverse conditions during training to improve the generalization capability of YOLOv7 for river sign detection in real-world environments.

4. Conclusion and Recommendation

4.1. Conclusion

Based on the results of the study conducted on the performance of the YOLOv7 model in detecting river signs under various image conditions, it can be concluded that:

1. The visual quality and conditions of the images have a significant impact on the model's accuracy, precision, and recall. Under original (normal) image conditions, YOLOv7 demonstrated its best performance with an accuracy of 78%, precision of 79.7%, and recall of 75.2%, indicating strong detection capability when the image quality is good and free from disturbances. However, the model's performance decreased significantly under degraded image conditions such as low resolution, noise, distortion (artifact), occlusion, and complex backgrounds (objects partially covered).
2. The most drastic performance decline occurred under occlusion and complex background conditions, where accuracy, precision, and recall ranged between 21–38%, showing that the model still has limitations in recognizing objects when parts of the signs are not visible or are covered by other elements. Meanwhile, blurred images were relatively better handled by the model, with detection performance remaining fairly stable.
3. Overall, this research demonstrates that YOLOv7 is effective for detecting river signs in good-quality images, but improvements are needed to enhance its generalization capabilities in handling complex environmental variations. Therefore, future development may be directed toward data augmentation, improved image preprocessing, or the use of hybrid models and specialized fine-tuning so that the detection system can operate reliably under real-world field conditions.

4.2. Recommendation

Based on the conclusions presented, the following recommendations can be suggested:

1. It is necessary to increase the quantity and variety of datasets with more diverse environmental conditions, such as extreme lighting, occlusion, and complex backgrounds, so that the model can better adapt to real-world field situations. The use of data augmentation techniques, such as rotation, contrast adjustment, and controlled noise addition, can also help improve the model's generalization capability.
2. Before the detection process, images should undergo enhancement procedures such as noise reduction, contrast improvement, or lighting adjustment. These steps can help clarify visual features on the signs so that the model can recognize objects more accurately.
3. Further training (fine-tuning) of the YOLOv7 model is required using river sign-specific data under various conditions so that the model weights align more closely with the characteristics of the target objects. Additionally, the use of transfer learning from other detection models may be considered to improve performance.
4. The developed model should be tested directly in real-world environments, such as river areas and docks, to evaluate the system's performance consistency against environmental variations that cannot be fully represented by laboratory datasets.

Acknowledgement

Thankyou for Politeknik Transportasi Sungai, Danau dan Penyeberangan Palembang and all parties involved in this research

References

[1] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. CVPR / arXiv.

[2] Ultralytics. YOLOv5 / Ultralytics YOLO documentation & repository. (GitHub / docs) —dokumentasi model YOLO modern dan panduan implementasi.

[3] Flores-Calero, M., Astudillo, C. A., Guevara, D., Maza, J., Lita, B. S., Defaz, B., ... Armingol Moreno, J. M. (2024). Traffic Sign Detection and Recognition Using YOLO Object Detection Algorithm: A Systematic Review. Mathematics, 12(2), 297. (Review yang merangkum banyak studi deteksi rambu/tanda menggunakan YOLO). McMillan, C., Zhao, J., Xue, B., Vennell, R., & Zhang, M. (2023). Improving Buoy Detection with Deep Transfer Learning for Mussel Farm Automation. arXiv:2308.09238. (Contoh penggunaan YOLO/transfer learning untuk pendeteksian buoy).

[4] KOLOMVERSE / KRISO (2022). KOLOMVERSE: Korea open large-scale image dataset for object detection in the maritime universe. (dataset besar untuk objek maritim termasuk buoy, lighthouse) — berguna sebagai referensi dataset maritim.

[5] Peraturan Menteri Perhubungan RI No. 25 Tahun 2011 tentang Sarana Bantu Navigasi Pelayaran (terkait warna & jenis rambu suar/pelampung).

[6] PM No. 52 Tahun 2012 tentang alur pelayaran sungai dan danau (ketentuan penyelenggaraan alur pelayaran pedalaman).

[7] Lee, S., Kim, H., & Park, J. (2023). Real-Time Traffic Sign Detection Using YOLOv7 for Autonomous Driving Systems. Sensors, 23(8), 4120.

[8] McMillan, C., Zhao, J., Xue, B., Vennell, R., & Zhang, M. (2023). Improving Buoy Detection with Deep Transfer Learning for Mussel Farm Automation. arXiv preprint arXiv:2308.09238.

[9] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[10] Ultralytics. (2024). YOLOv7 Documentation. Retrieved from https://docs.ultralytics.com

[11] Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv preprint arXiv:2207.02696.

[12] Azmi, M. A., Rahman, M. N., & Supriadi, A. (2023). Penerapan Teknologi Computer Vision untuk Deteksi Objek Navigasi Sungai Menggunakan Deep Learning. Jurnal Teknologi Transportasi Air, 5(2), 45–54.

[13] Kementerian Perhubungan Republik Indonesia. (2011). Peraturan Menteri Perhubungan Nomor PM 25 Tahun 2011 tentang Sarana Bantu Navigasi Pelayaran.

[14] Kementerian Perhubungan Republik Indonesia. (2012). Peraturan Menteri Perhubungan Nomor PM 52 Tahun 2012 tentang Alur Pelayaran Sungai dan Danau.

[15] Rachman, A., Widodo, D. S., & Prasetyo, H. (2020). Analisis Efektivitas Rambu Sungai dalam Mendukung Keselamatan Pelayaran di Perairan Pedalaman. Jurnal Transportasi Perairan Indonesia, 8(1), 23–31.

[16] Azmi, M. A., Rahman, M. N., & Supriadi, A. (2023). Penerapan Teknologi Computer Vision untuk Deteksi Objek Navigasi Sungai Menggunakan Deep Learning. Jurnal Teknologi Transportasi Air, 5(2), 45–54.

[17] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

[18] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444.

[19] Szeliski, R. (2022). Computer Vision: Algorithms and Applications (2nd ed.). Springer.

[20] Zhao, Z. Q., Zheng, P., Xu, S. T., & Wu, X. (2019). Object detection with deep learning: A review. IEEE Transactions on Neural Networks and Learning Systems, 30(11), 3212–3232.