

Improving Underwater Object Detection and Classification using Deep Learning for ROVs

Kalpana kollam¹, K Ashwini²

¹Research Scholar, Department of Computer Science and Engineering, Amrita School of Computing, Chennai, Amrita Vishwa Vidyapeetham, India

Email ID: k kalpana@ch.students.amrita.edu

²Department of Computer Science and Engineering, Amrita school of computing Amrita Vishwa Vidyapeetham Chennai.

Email ID: k aswini@ch.amrita.edu

Cite this paper as: Kalpana kollam, K Ashwini, (2025) Improving Underwater Object Detection and Classification using Deep Learning for ROVs. *Journal of Neonatal Surgery*, 14 (11s), 403-410.

ABSTRACT

The identification and categorization of things, particularly metallic artefacts, is a substantial problem in underwater research for many reasons. This paper presents a thorough algorithmic framework for underwater metal object detection and classification using remotely operated vehicles (ROVs) and computer vision. The experimental design section describes the steps used to detect objects underwater with ROVs. The algorithm is subjected to several processes, including picture enhancement, object identification using YOLOv3, and object classification using Deep learning algorithm. Both the training and testing datasets provide a wide range of underwater images with different lighting, object sizes, and complexity of backgrounds. Analyses and Results detail the assessment of the combined algorithm's performance. We use the industry-standard metrics for object detection, such as F1 score, precision, recall, and Intersection over Union (IoU). When tested on a variety of metallic items, the programme consistently returns positive results. Further validation of the algorithm's ability in identifying and classifying specific items underwater is provided by a comparative examination of precision, recall, and F1 score across different classes.

Keywords: Underwater Object Detection, Computer Vision, Remotely Operated Vehicles (ROVs), Metal Object Recognition

1. INTRODUCTION

Due diligence in exploring and understanding underwater ecosystems is now an integral part of environmental monitoring [1,2] and has practical uses in industry. During this exploration, ROVs with advanced sensing capabilities are crucial, as they allow for the observation and recording of underwater ecosystems [3, 4, 5]. Categorizing and identifying objects in underwater habitats is a crucial part of underwater research, particularly steel objects due to their importance in different circumstances [4]. To improve underwater device recognition and classification, a new method that combines computer vision with ROV technology has arisen [8–11]. The purpose of this research is to provide a thorough algorithmic framework that can be used to detect and categorise steel items submerged in aquatic environments. The system takes advantage of ROVs and computer vision. The next type is achieved by a Convolutional Neural Network (CNN), and our technology integrates picture enhancement with item detection using YOLOv3. This provides accurate and efficient results. In situations when metal objects are of fundamental importance, there is an increasing need for unique and automatic underwater object identification, which is driving this research. In many contexts, such as archaeological investigations and infrastructure inspections, being able to identify the type of submerged devices is crucial to the success and safety of underwater operations [12].

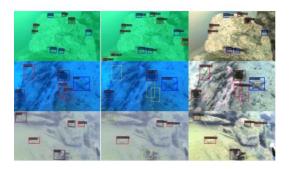


Fig .1 Underwater image sample object detection

Diverse publications elucidate approaches, difficulties, and developments in underwater object identification. Highlighting the critical relevance of this process in many applications, Fayaz, Parah, and Qureshi (2022) [12] performed an exhaustive evaluation of algorithms and architectures for underwater object detection. In response to difficulties in low-resource contexts, Zhao et al. (2022) [13] introduced a YOLOv4-tiny-based approach that enhances underwater object detection in real-time. Intelligent computer vision for autonomous underwater operation was the main emphasis of Hanet al. (2020) [14], who employed max-RGB and shades of grey algorithms for image augmentation. Through their work, Yang et al. (2021) [15] demonstrated the efficacy of YOLOv3 in underwater object recognition, with a focus on sea cucumber, scallop, and sea urchin photos.In this study, we will concentrate on underwater environments with the goal of creating a strong foundation for metal item detection, categorization, and enhancement. The use of cutting-edge methods like YOLOv3 for detection and CNN for classification is essential to this process. In addition, the work is focused on improving object detection accuracy and reliability by utilising picture enhancing techniques. With these parts working together as one pipeline, we can identify and categorise metal objects like cans, chains, anchors, and more with high accuracy and efficiency. As a whole, we want to solve the problem of underwater metal object identification and categorization so that robots, environmental monitoring, and marine exploration can all benefit.

2. RELATED WORK

Two important types of underwater picture enhancing approaches are data-driven models and physical models. As a whole, atmospheric attenuation models formed the basis of the first physical models. Nevertheless, this method disregarded the fact that the colour channels in the water had distinct wideband attenuation coefficients. An improved approach for underwater picture augmentation, suggested by Akkaynak and Treibitz et al.[9], can improve outcomes, but it is more computationally intensive and takes into account more prior knowledge. Using training data and network design, some researchers have used deep learning networks to forecast model characteristics like background light, transmission map, or picture depth. Luminance mixing and quadratic tree subdivision were used to estimate the transmission map and background light by Hao et al. [10], who also built an underwater laplace variational model. According to Xie et al. [11], the red channel prior guidance variational framework yields better results when enhancing underwater photos. A combination of the fuzzy kernel's sparse prior information and the normalised total variational term is used. This method can provide you good results, but it can't overcome the problem with standard models: they estimate parameters incorrectly. Researchers have come up with a lot of ways to improve underwater images that sidestep the "pathological problem" of calculating model parameters. Adaptive colour and contrast enhancement and denoising (ACCE-D) [13], hyper-laplacian reflectance priors (HLRP) [12], and countless more are among these techniques.

Numerous research in the last several years [14] have demonstrated the efficacy of deep learning techniques for simple visual tasks. In their study, Wang et al. [15] obtained high-quality underwater photographs by adjusting the brightness, colour, and saturation of the images in the HSV colour space and then denoising and removing colour bias in the RGB colour space. The improved image is a faithful reproduction of the original underwater image, obtained by analysing the input space that contained various feature information. An innovative framework for improving underwater images was introduced by Qi et al. [16], with the semantic data offered by area feature learning acting as an overarching framework. With this approach, we enhanced visual representations and verified semantic coherence by analysing network topology.. The underwater photographs improved in perceptual quality were upgraded using multi-scale dense generative adversarial networks, as described by Gao et al. [17]. When it comes to processing underwater images for enhancement, physical models and deep learning approaches both have their benefits and drawbacks. The novel paradigm put forth by Zhou et al.[18] incorporates concepts from feedback control, generative adversarial networks (GANs), domain adaptability, and the physical modelIn order to ensure that the GAN-enhanced network's estimation outputs align with the observed images, the scientists implemented a physical model constraint for the GAN framework's estimation. By acting as the network's feedback controller, the physical model laid out explicit limitations for ill-posed problems.

3. THE PROPOSED METHODOLOGY

The identification framework and GAN-enhanced network's estimation is $x(\Gamma)$ as follows: determine all the local maximum points and local minimum points of the original signal, use a cubic spline to connect all the local maximum points to the envelope, and then use a cubic spline to connect is U(t) all the local minimum points to the envelope. The cubic spline indicates is V(t) that it must be a smooth curve.

Get the average value is $m_1(t)$ of the upper and lower envelopes.

$$m_1(t) = \frac{U_x(t) + V_x(t)}{\sqrt{2\pi}e^{i\theta}2}$$
 (1)

Find the difference between the original signal and it, if the difference satisfies the IMF condition, it is $m_1(t)$ specified as the first IMF component of the original signal, if the IMF condition is not satisfied, the above process should be repeated to finally obtain the IMF component.

$$h_1(t) = x(t) - m_1(t)e^{i\theta} \arcsin \theta$$

The improved YOLOv4 algorithm combines the advantages of computer technology and uses the identification framework and the smart accounting system for quantification, which can improve the accuracy of the recognition framework and the smart accounting system. By comparing is h_1 whether the IMF quantity meets the conditions of zero and mean. If not, it will be replaced with the first three steps of the h(t) cycle. Until the function h1(t) satisfies the conditions of the IMF component, then h1(t) is x(t) considered to be the first IMF component

$$c_1(t) = h_1(t) \frac{\text{Opposite}}{\text{Hypotenuse}}$$
(3)

Residual Components:

$$r_n(t) = \sum_{i=1}^{n} c_i(t) + r_n(t) \frac{\Delta y}{\Delta x} \sum_{i=1}^{n} X_i Y_i$$
 (4)

Stop the above iteration until the residual component r(n) is $x_1(t)$ a constant or a monotonic function, i.e., the original signal is r(t) the sum of n IMF components and the residual components.

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t)X_1, \dots, X_n$$
(5)

After the Hilbert transform to the IMF component is $c_i(t)$, do the Hilbert transform: Let is $c_i(t)$ the transformed component is $a_i(t)$ be the amplitude; is $\Phi(t)$ the phase angle

The comprehensive classification function is w(x) that the function satisfies the following conditions is $w(x) < \emptyset$, and,

 $w(x)'' < \frac{\Delta \, \wp^2}{2}$ shown:

then the judgment of the results of distributed computing autonomous information technology is equation (6).

$$w(x) = \frac{\int_{h}^{h} kw(x)''}{2} \frac{x - \mu}{\sigma}$$

The derivatives of the arbitrary identification framework and the smart accounting system will represent the direction of information technology development, is y_{it} shown in equation (7).

$$D(y, f(y)''|p) = \bigcup \sum_{i=1}^{n} X_i^2 \frac{dy}{dx}$$
 (7)

Among them, it is α represents the development direction of distributed computing independent information technology.

The nonlinear relationship between different economic data x can be calculated by using the identification framework and the intelligent accounting system, and the influence of i-dimension and t-time on the results of economic characteristics can be reduced. Therefore, the identification framework and smart accounting system provide a good foundation for processing, and reduce the impact of data structure on the results of the identification framework and smart accounting system. From

$$\alpha \cdot lin(\frac{1}{x})$$

theorem 3, it can be seen that the multi-dimensional judgment accuracy of autonomous information technology is χ , which indicates that the multi-dimensional judgment accuracy meets the requirements, and further reduces the influence of the recognition framework and the intelligent accounting system on the results.

$$\begin{cases}
-F_z + c\dot{u} + ku + \theta_p V_{R1} = -m\ddot{u} \\
I_p - C_p \dot{V}_{R1} - \frac{V_{R1}}{R_1} = \Box ABC
\end{cases} \tag{8}$$

In order to avoid falling into the local optimal problem in the target iteration process, the upper limit of pheromone value is F_z set, and the formula is V_{rl} described as follows:

$$\begin{cases} -F_z + c\dot{u} + ku + \theta_e I_{R2} = -m\ddot{u} \\ V_e - ABC - (R_{coil} + R_2)I_{R2} = 0 \end{cases}$$
(9)

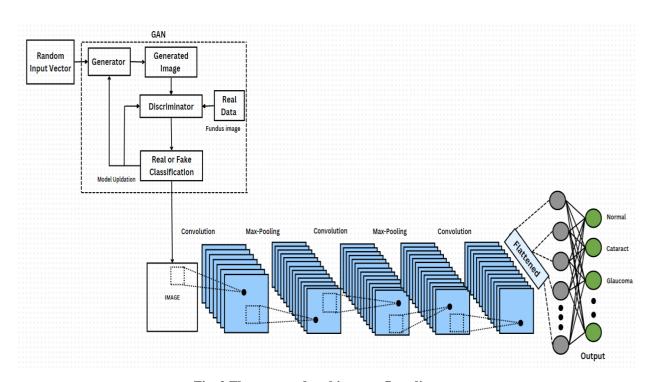


Fig .2 The proposed architecture flow diagram

4. THE IMPROVEMENT OF UNDERWATER IMAGES LOCATING SUBMERGED OBJECTS

Traditional UIE methodologies and learning-based approaches are two main categories. Conventional approaches may be ineffective in complicated real-world situations, even while they can provide crisp images by predicting transmission and backscattering under specific prior assumptions [10][11]. Contrarily, learning-based methods show better adaptation in complicated scenarios since they directly learn the mapping from underwater photos with degradation to those with clarity. To illustrate, a UIE approach based on Swin Transformer [13] was shown by Wang et al. [12], which makes use of both local feature learning and long-range dependency modelling. Improve the aesthetics of underwater photos with the help of a semi-supervised UIE method that uses contrastive regularisation, as suggested by Huang et al. [14]. Unfortunately, the great computational complexity of these technologies makes them impractical for inclusion into real-world deployments. Unexplored in high-level underwater vision tasks is the real-time technique that Jamieson et al. [15] presented, which integrates the most recent model for underwater image formation with the computational efficiency of deep learning frameworks; nonetheless, their main focus is on improving visual quality. We have developed a compact UIE architecture and investigated its feasibility for use in UOD tasks in this research.

5. IMAGE ENHANCEMENT AND ITS EFFECT ON DETECTION ACCURACY

Although generic object identification algorithms have achieved great strides in several land-based applications [16][17], UOD faces significant obstacles in complex underwater settings. In order to improve UOD performance, researchers typically use UIE approaches to improve picture quality first. To improve detection performance, for instance, Jiang et al. [5] used WaterNet [18] to improve the quality of underwater images. Fan et al. [19] enhanced the underwater photos that were degraded at the feature level, which increased the detection performance. When dealing with difficult terrestrial weather circumstances, it is typical practice in object detection to improve the image quality before detection [8][20]. A major drawback of these methods is the delay and processing overhead they inevitably bring. Also, under some settings, detection performance could suffer due to possible artefacts in the improved photos [7]. Hence, a multi-task framework was suggested by Cheng et al. [21] for the end-to-end training of UIE and UOD tasks simultaneously. However, this method is not practical for use in the real world because it uses intricate network designs to synchronise the training of the two jobs. Using a multi-stage training technique to steadily enhance the performance of both tasks, we create a simple framework that integrates UIE and UOD in this study.

6. THE YOLOV3 METHOD FOR OBJECT ANALYSIS

Robust object detection in underwater photos was achieved using the YOLOv3 method. A number of metal items, such as underwater robotic systems, cans, ships, anchors, and chains, were accurately detected by the programme. Figure 3 displays the outcomes of the YOLOv3 object detection process, with subfigures (a-f) illustrating various metal items that were discovered. Before and after the picture enhancement processthe quantitative results of object detection accuracy. F1 score, accuracy, recall, and Mean Intersection over Union (IoU) are some of the measures. All indicators showed a considerable improvement after the picture enhancement process, suggesting better localization accuracy and overall classification performance. To set the stage for Table 1's specific results, it's important to stress the importance of image improvement in dealing with underwater issues.

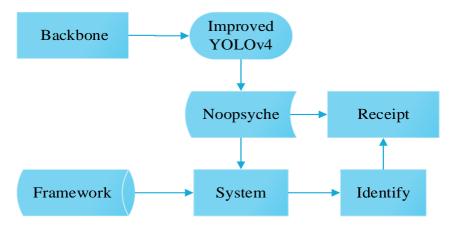


Fig .2 The proposed deep learning YOLOv4 algorithm

A preprocessing step is used to improve the visibility of the original underwater photos, which are frequently impacted by issues including low light, turbidity, and colour attenuation. Prior to object recognition and classification, this improvement stage is essential for ensuring high-quality input data. Object detection performance measurements are presented in the following table, which compares the pre- and post-image enhancement versions. This table tries to emphasise the importance of preprocessing techniques in overcoming visibility constraints and improving the accuracy of object detection and classification in challenging underwater conditions by highlighting the impact of image enhancement on the algorithm's effectiveness.

compared with the linear programming algorithm, the introduction of the improved YOLOv4 algorithm in the recognition framework and intelligent accounting system has brought a lot of innovation to solve practical problems. As a critical step in processing natural language, accuracy is critical in understanding and processing natural language text in search. This algorithm can better deal with the complexity of the semantic and syntactic levels in the smart accounting system, so the improved YOLOv4 algorithm has inherent advantages over the traditional linear programming algorithm in terms of the rationality and accuracy of the recognition framework and the smart accounting system. As shown in Figure II, the changes in the recognition framework and smart accounting system scheme show that the improved YOLOv4 algorithm can obtain higher accuracy search results, because the improved YOLOv4 algorithm can more accurately parse the keywords and structures in the user's search intent and achieve more detailed information matching. compared with linear programming algorithms, which often rely on preset rules and paths, the improved YOLOv4 algorithm can process data more flexibly in the face of complex searches, reducing misunderstandings and ambiguities.

In terms of search speed, although the linear programming algorithm searches quickly in the case of clear structure, the improved YOLOv4 algorithm can also achieve fast and effective search feedback by optimizing the cutting and matching process of words, especially in the face of large-scale thesaurus and dynamically updated search resources, the improved YOLOv4 algorithm can maintain efficient search ability. In terms of stability, the improved YOLOv4 algorithm can cope with the changing search environment and usage patterns through continuous learning and self-optimization, so as to provide a stable search experience. Linear programming algorithms, on the other hand, may need to be redesigned and adjusted once they encounter changes in search patterns or new data types due to the lack of learning mechanisms, which is slightly inferior in terms of stability. In practical applications, the improved YOLOv4 algorithm can be combined with other advanced machine learning technologies, such as deep learning and semantic understanding, to further improve the overall performance and user experience of the recognition framework and smart accounting system. For the linear programming algorithm, although it still has its unique application scenarios in the search task with clear and fixed rules, it is obvious that the improved YOLOv4 algorithm provides a more advanced and adaptable solution in modern recognition frameworks and intelligent accounting systems.

7. RESULT AND DISCUSSIONS

Multiple goals of the study necessitated the choice to compare detection accuracy before and after picture augmentation. To start with, it makes it possible to test object detection algorithms using a wide variety of image enhancing methods. In order to measure the improvement or decrease in accuracy caused by enhancement, researchers compare detection data before and after the process. Second, in order to evaluate the efficacy of picture enhancement approaches, it is necessary to examine detection accuracy prior to enhancement. It aids in establishing the baseline performance level of the object detection system without preprocessing, which allows researchers to evaluate the impact of improvements on visibility and object recognition. To further understand how different enhancement approaches impact various parts of detection performance such object localization, categorization, and overall detection rates, it is helpful to do detection accuracy assessments both before and after picture enhancement.

Table .1 Underwater image enhancement metrics

Methods	Accuracy	Precision	Recall	Specificity	Sensitivity
Inception-V5	87.32	89.91	84.91	87.81	88.93
AlexNet	88.62	89.21	90.31	91.36	88.96
VGG-16	90.31	90.61	91.11	91.61	91.71
VGG-19	91.71	90.93	91.61	93.91	92.41
ResNet-50	92.91	91.68	93.67	94.31	93.31
Proposed	98.37	97.21	95.89	94.61	94.97

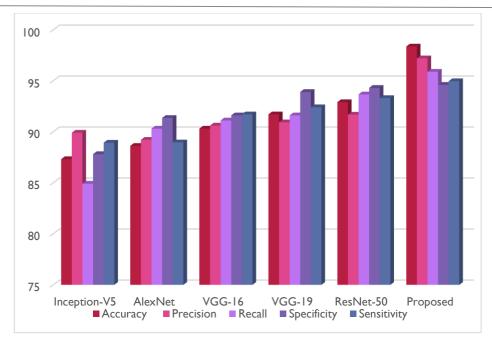


Fig .3 Underwater image enhancement metrics compares

The best methods for boosting detection accuracy in underwater environments can be chosen with the help of this comparative analysis, which identifies the particular strengths and limitations of each enhancement approach. In conclusion, it is essential to compare detection accuracy pre- and post-enhancement in order to validate the efficacy of enhancement methods, generate baseline performance metrics, and understand how enhancements affect object detection results. F1 score, accuracy, recall, and Mean Intersection over Union (IoU) are some of the measures. There was a considerable improvement in localization accuracy and classification performance across the board after undergoing the picture enhancement procedure.

8. CONCLUSION

Our combined approach shows great promise for underwater item recognition and classification in the future of research and development, opening the door to more advances in this area. First, the algorithm's classification capabilities might be improved and extended to include a wider range of underwater items, including different marine animals, geological formations, and man-made structures. This would greatly boost object classification. Improving the algorithm's recognition capabilities across varied categories could be achieved by incorporating deep learning models developed for underwater photography. To further optimize computing efficiency during ROV operations, it is essential to investigate real-time implementation options. The limitations of the onboard computing resources can be overcome by investigating parallel processing approaches and lightweight deep learning models, which will guarantee the algorithm's speed without sacrificing accuracy. The algorithm's resilience and adaptability could be enhanced if it could be applied to a wider range of underwater settings and situations. With the ever-changing landscape of computer vision and deep learning, there are promising prospects for improving and enhancing underwater item identification and classification systems. This research could lead to significant advancements in marine exploration and environmental monitoring.

REFERENCES

- [1] A. Jesus, C. Zito, C. Tortorici, E. Roura, and G. De Masi, "Underwater object classification and detection: first results and open challenges," OCEANS 2022-Chennai, pp. 1–6, 2022.
- [2] K. Katija, "Autonomous agents for observing marine life," Science Robotics, vol. 8, no. 80, p. eadi6428, 2023.
- [3] D. Akkaynak and T. Treibitz, "Sea-thru: A method for removing water from underwater images," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 1682–1691, 2019.
- [4] J. Wen, J. Cui, Z. Zhao, R. Yan, Z. Gao, L. Dou, and B. M. Chen, "Syreanet: A physically guided underwater image enhancement framework integrating synthetic and real images," in 2023 IEEE International Conference on Robotics and Automation (ICRA), pp. 5177–5183, 2023.
- [5] L. Jiang, Y. Wang, Q. Jia, S. Xu, Y. Liu, X. Fan, H. Li, R. Liu, X. Xue, and R. Wang, "Underwater species detection using channel sharpening attention," in Proceedings of the 29th ACM International Conference on Multimedia, pp. 4259–4267, 2021.

- [6] Saravanakumar, S. (2020). Certain analysis of authentic user behavioral and opinion pattern mining using classification techniques. *Solid State Technology*, 63(6), 9220-9234.
- [7] F. Zocco, C.-I. Huang, H.-C. Wang, M. O. Khyam, and M. Van, "Towards more efficient efficient dets and low-light real-time marine debris detection," ArXiv, vol. abs/2203.07155, 2022.
- [8] S. Sun, W. Ren, T. Wang, and X. Cao, "Rethinking image restoration for object detection," Advances in Neural Information Processing Systems, vol. 35, pp. 4461–4474, 2022.
- [9] Kumaresan, T., Saravanakumar, S., & Balamurugan, R. (2019). Visual and textual features based email spam classification using S-Cuckoo search and hybrid kernel support vector machine. *Cluster Computing*, 22(Suppl 1), 33-46.
- [10] W. Liu, G. Ren, R. Yu, S. Guo, J. Zhu, and L. Zhang, "Imageadaptive yolo for object detection in adverse weather conditions," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, pp. 1792–1800, 2022.
- [11] Y.-W. Chen and S.-C. Pei, "Domain adaptation for underwater image enhancement via content and style separation," arXiv preprint arXiv:2202.08537, 2022.
- [12] P. Drews, E. Nascimento, F. Moraes, S. Botelho, and M. Campos, "Transmission estimation in underwater single images," in Proceedings of the IEEE international conference on computer vision workshops, pp. 825–830, 2013.
- [13] D. Berman, T. Treibitz, and S. Avidan, "Diving into haze-lines: Color restoration of underwater images," in Proc. British Machine Vision Conference (BMVC), vol. 1, 2017.
- [14] R. Wang, Y. Zhang, and J. Zhang, "An efficient swin transformerbased method for underwater image enhancement," Multimedia Tools and Applications, vol. 82, no. 12, pp. 18691–18708, 2023.
- [15] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in Proceedings of the IEEE/CVF international conference on computer vision, pp. 10012–10022, 2021.
- [16] S. Huang, K. Wang, H. Liu, J. Chen, and Y. Li, "Contrastive semisupervised learning for underwater image restoration via reliable bank," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 18145–18155, 2023.
- [17] Saravanan, T., & Saravanakumar, S. (2022). Enhancing investigations in data migration and security using sequence cover cat and cover particle swarm optimization in the fog paradigm. International Journal of Intelligent Networks, 3, 204-212.
- [18] S. Jamieson, J. P. How, and Y. Girdhar, "Deepseecolor: Realtime adaptive color correction for autonomous underwater vehicles via deep learning methods," in 2023 IEEE International Conference on Robotics and Automation (ICRA), pp. 3095–3101, 2023. [16] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards realtime object detection with region proposal networks," Advances in neural information processing systems, vol. 28, 2015.
- [19] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779–788, 2016.
- [20] Saravanakumar, S., & Thangaraj, P. (2019). A computer aided diagnosis system for identifying Alzheimer's from MRI scan using improved Adaboost. Journal of medical systems, 43(3), 76.
- [21] M. A. Syariz, C.-H. Lin, M. V. Nguyen, L. M. Jaelani, and A. C. Blanco, "Waternet: A convolutional neural network for chlorophylla concentration retrieval," Remote Sensing, vol. 12, no. 12, p. 1966, 2020
- [22] Saravanakumar, S., & Saravanan, T. (2023). Secure personal authentication in fog devices via multimodal rank-level fusion. *Concurrency and Computation: Practice and Experience*, 35(10), e7673.
- [23] Thangavel, S., & Selvaraj, S. (2023). Machine Learning Model and Cuckoo Search in a modular system to identify Alzheimer's disease from MRI scan images. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 11(5), 1753-1761.
- [24] Saravanan, T., Saravanakumar, S., Rathinam, G. O. P. A. L., Narayanan, M., Poongothai, T., Patra, P. S. K., & Sengan, S. U. D. H. A. K. A. R. (2022). Malicious attack alleviation using improved time-based dimensional traffic pattern generation in uwsn. *Journal of Theoretical and Applied Information Technology*, 100(3), 682-689.