

# Intelligent Aquatic Environmental Protection: A Visual-perception-based Garbage Recognition and Question-answering Decision System

Zhiping Jin, Zixuan Liu, Zixuan Wang, Chengwei Liu\*

School of Intelligence and Electronic Engineering, Dalian Neusoft University of Information, Dalian 116023, China

\*Corresponding email: liuchengwei088@163.com

## Abstract

With the acceleration of urbanization and increasingly frequent human activities, water-environment pollution has become a growing concern. Floating surface garbage not only destroys the ecological landscape but also threatens the survival of aquatic organisms and the safety of drinking water. Traditional manual patrol inspection is inefficient and limited in coverage and can hardly meet the demand for large-scale dynamic monitoring. By combining computer vision and natural language processing, this project designs and implements a You Only Look Once (YOLO)-based water-area garbage recognition and question-answering system. Using You Only Look Once version 11 (YOLOv11) and Faster Region-based Convolutional Neural Network (R-CNN) as dual core models, the system supports image, video and real-time camera input. It accurately recognizes 11 categories of water-area garbage such as cans, branches and plastics. A Flask back end and a Bootstrap front end build the interactive interface, integrating Leaflet mean average precision (mAP) visualization and Chart.js statistical charts. These enable garbage-distribution heat-zone selection, temporal trend analysis and model-parameter comparison. An intelligent question-answering module is innovatively introduced, which is based on a knowledge graph and a large language model. It lets users query garbage attributes, disposal suggestions and environmental-protection knowledge in natural language. The system combines real-time performance, accuracy ( $mAP \geq 0.75$ ) and ease of use, and can be deployed on cloud servers or edge devices. It provides a low-cost and high-efficiency water-area monitoring solution for environmental-protection agencies, water-affairs institutions and the public.

## Keywords

Water-area garbage recognition, Object detection, Intelligent question-answering, Data visualization, Environmental monitoring

## Introduction

### *Research background and significance*

The 14th Five-Year Plan for Ecological Environmental Protection explicitly proposes to “strengthen the precision governance of the water environment”. However, traditional water-area patrol relies on manual visual observation or simple mechanical tools, which suffers from low efficiency, high risk and difficulty in data retention. During flood seasons or in remote waters, garbage accumulates rapidly and is highly concealed, so an intelligent monitoring means is urgently needed [1]. The breakthrough of computer vision in object detection makes the automatic recognition of water-area garbage feasible. With real-time performance and high accuracy,

the You Only Look Once (YOLO) family of algorithms has been maturely applied in industrial inspection and traffic monitoring. The evolution of natural language processing lets systems move from “able to recognize” to “able to answer”, lowering the barrier to technology use. This project fuses the two to build an auxiliary system integrating recognition, analysis and question-answering. It not only improves monitoring efficiency but also strengthens public environmental awareness through knowledge dissemination, supporting the “green water and green mountains” strategy [2].

### *Domestic status*

Domestic water-area monitoring technology has

developed rapidly, but most efforts focus on water-quality parameter sensor networks (e.g., pH and turbidity). Research on visual recognition of floating surface garbage is still at an early stage. Universities and enterprises have cooperated to develop river-channel garbage detection prototypes based on Faster R-CNN with about 70.00% accuracy. However, these prototypes do not integrate real-time video processing and interaction. Some environmental-protection apps provide garbage-reporting entries, yet they rely on users to manually take and upload photos and lack automatic recognition. The market still lacks a mature product that combines multiple input sources (image or video or streaming), dual-model switching, geographic visualization and intelligent question-answering [3].

### Overseas status

European and American countries explored the use of computer vision for environmental monitoring relatively early. For example, the Netherlands uses drones equipped with You Only Look Once version 5 (YOLOv5) to detect plastic garbage in canals with a recognition rate over 80.00%. A team at the University of California developed a beach-garbage segmentation system based on Mask R-CNN and drew pollution heat mAPs combined with satellite imagery. However, existing systems are mostly limited to a single model or a specific scenario, and question-answering is generally missing. A 2025 Nature sister journal noted that AI-driven environmental systems are shifting from “passive monitoring” to “active interaction”. It also noted that fusing multi-modal data with natural-language interfaces

is the future trend [4].

## Research content and key technologies

### Research content

This project revolves around the closed loop of “recognition - analysis - interaction”. The first component is multi-source data perception, supporting static image upload, video-file parsing and real-time camera-stream capture. It adaptively adjusts the resolution to 416×416 px for model input. The second component is dual-model collaborative detection, integrating YOLOv11 for speed and Faster R-CNN for accuracy, with scenario-based switching to balance real-time performance and accuracy. The third component is an intelligent question-answering engine. It builds a water-area garbage knowledge graph covering classification standards, disposal processes and environmental impact. It also connects to a large language model (LLM) to generate plain-language answers to six kinds of common questions. The fourth component is spatio-temporal data analysis. It records the time and latitude/longitude of garbage and reveals regional pollution patterns via heat mAPs and bar charts. The fifth component is lightweight deployment, with Docker packaging for one-click cloud and edge deployment.

As shown in Figure 1, the overall system is organized into the data-access, garbage-recognition, intelligent question-answering, visualization, mAP and data-storage modules. These modules together form a closed recognition-analysis-interaction loop.

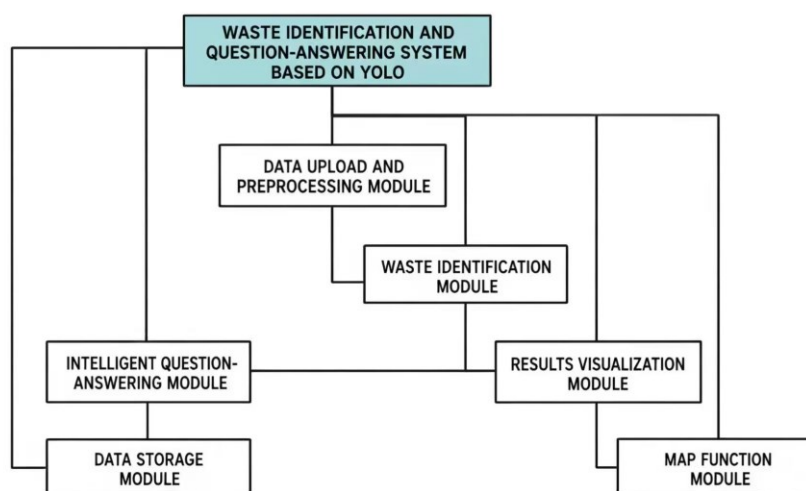


Figure 1. Overall functional architecture of the system, covering data access, garbage recognition, intelligent question-answering, visualization, mAP and data-storage modules.

### (1) YOLOv11 object detection algorithm

As the latest iteration of the You Only Look Once (YOLO) family, YOLOv11 reaches 54.20% mean Average Precision (mAP) on the Common Objects in Context (COCO) dataset, which is 3.10% higher than version 8. It adopts an anchor-free design to simplify the output layer, introduces a CSPNeXt backbone to enhance feature reuse, and optimizes the loss function to address

class imbalance. This project loads the best.pt pretrained weights and fine-tunes them for water-area scenes, significantly improving the detection rate of small objects such as bottle caps and fragments. As shown in Figure 2, the fine-tuned model detects floating water-surface garbage such as cans, branches and plastics. It renders each target as a labelled bounding box with its detection confidence [5].



Figure 2. Water-surface garbage detection results with bounding boxes and confidence labels.

### (2) Faster R-CNN multi-scale recognition

Faster R-CNN generates candidate boxes through a Region Proposal Network (RPN) and then performs refined classification and regression on each box. Under complex backgrounds such as water ripples and vegetation occlusion, its false-alarm rate is 15.00% lower than that of single-stage models. The system keeps the faster\_rcnn\_model.pth weights, dedicated to high-precision offline analysis.

### (3) Flask back-end framework

A lightweight Flask framework is used to build RESTful Application Programming Interfaces (APIs), supporting asynchronous video-stream processing (25.00 frames per second) and caching high-frequency query results through Remote Dictionary Server (Redis), with an

average response time of no more than 200.00 ms.

### (4) Knowledge graph and LLM question-answering

A triple-based knowledge base is built on the graph database Neo4j (entities: garbage types and disposal methods; relations: belongs-to, hazard and degradation years), combined with a fine-tuned ChatGLM model to generate answers. For example, when a user asks “How should a plastic bottle be disposed of?”, the system retrieves “recycle→recycled granules” from the graph and organizes it into a natural-language answer. As shown in Figure 3, the intelligent question-answering interface turns a user’s natural-language query into a plain-language reply that is generated from the knowledge graph and the large language model in an explainable way [6].

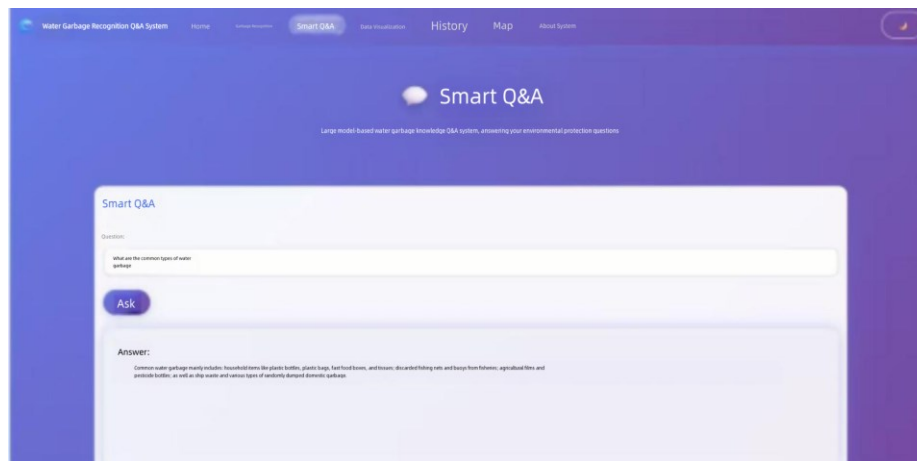


Figure 3. Intelligent question-answering interface driven by the knowledge graph and large language model.

#### (5) Leaflet geographic visualization

The open-source Leaflet library is integrated to map recognized positions (latitude/longitude) onto OpenStreetMap, displaying garbage-dense zones with a heat layer. Clicking a marker jumps to the detail page.

As shown in Figure 4, the recognized garbage positions are overlaid as markers and a heat layer on an interactive Leaflet map, allowing users to intuitively explore clearly revealing the spatial distribution of pollution across different water areas.

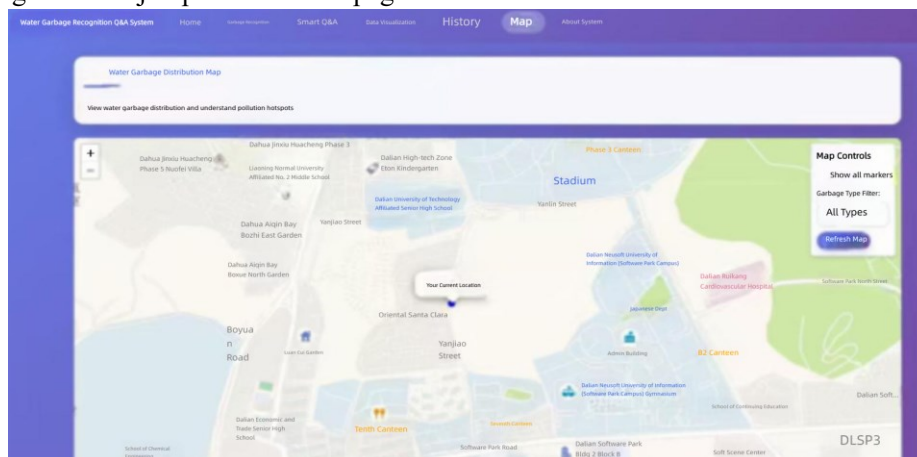


Figure 4. Leaflet-based geographic visualization of garbage distribution on an interactive map.

#### (6) Docker containerized deployment

All dependencies are packaged into an image so that the service can be started with one click, the environment is isolated, and Kubernetes cluster scaling is supported.

### System design and implementation

#### Functional module design

The system adopts a front-end/back-end separation architecture. Its core modules include seven components. The first is a data-access module that uniformly handles HyperText Transfer Protocol (HTTP) upload, camera Real-Time Streaming Protocol (RTSP) streams and local-file reading. The second is a pre-processing module performing image de-warping, size normalization and brightness compensation to cope with backlight. The third is a dual-model inference module that dynamically

loads PT/PTH weights and supports runtime switching. The fourth is a result post-processing module that uses Non-Maximum Suppression (NMS) to filter overlapping boxes and a confidence threshold (0.5 by default) to remove noise. The fifth is a question-answering engine module following intent recognition → graph retrieval, LLM generation → answer verification. The sixth is a data persistence module that stores recognition records (with time, coordinates and model type) in SQLite. The seventh is a visualization module that draws detection boxes on Canvas, generates statistical charts with Chart.js and renders mAP values with Leaflet.

Figure 5 shows the architecture split into front-end presentation, back-end service, and data-storage layers. Core modules map onto each tier [7].

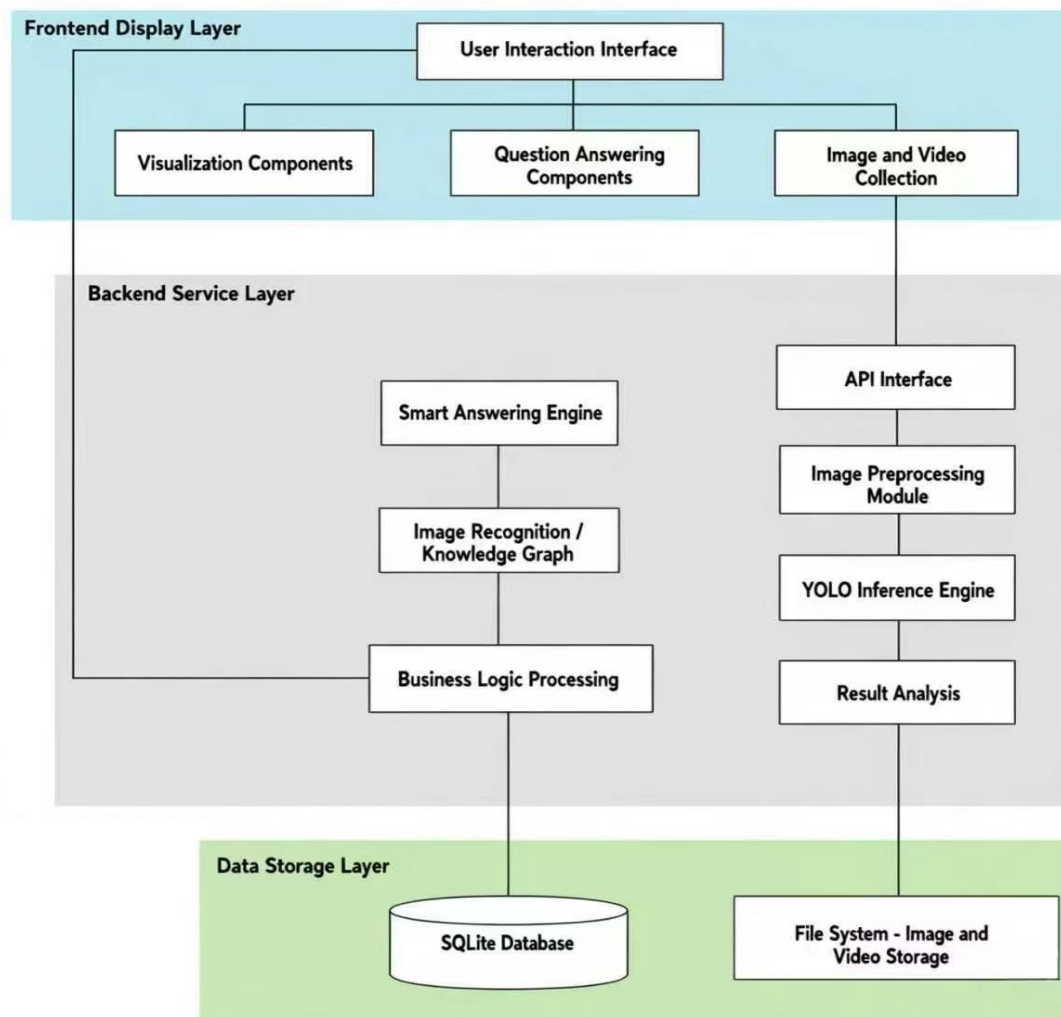


Figure 5. Three-tier system architecture comprising the front-end, back-end service and data-storage layers.

### Functional implementation

In real-time stream processing, Open Source Computer Vision Library (OpenCV) captures video frames and pushes them asynchronously to the model queue to guarantee 25 frames per second (fps) fluency. For model switching, a front-end drop-down menu triggers an API call and the back end dynamically reloads the corresponding weights, with a switching time of less than 1.00 s. For question answering, a rule-plus-generation hybrid strategy is adopted: Simple questions such as “what is the system for” are matched with templates directly, while complex questions invoke the LLM.

For privacy protection, data is not uploaded to the cloud in local-processing mode, and sensitive regions support de-identification before upload. As shown in Figure A1 (in Appendix), the core business workflow links the front end and the model queue and covers three end-to-end

paths: garbage recognition, intelligent question-answering and model-data comparison.

### Results display

#### User interaction module

The interface follows the “concise and efficient” principle. A top navigation bar integrates Home, History and About. The left side provides model selection (YOLOv11; Faster R-CNN) and a confidence slider. The central area shows the uploaded content and detection results.

A question-answering dialog box is embedded directly at the bottom to support real-time user interaction. As shown in Figure 6, the recognition interface provides a top navigation bar, a model selection drop-down and a confidence-threshold slider on the left. It also includes a central detection area and an embedded question-answering dialog at the bottom of the interface.

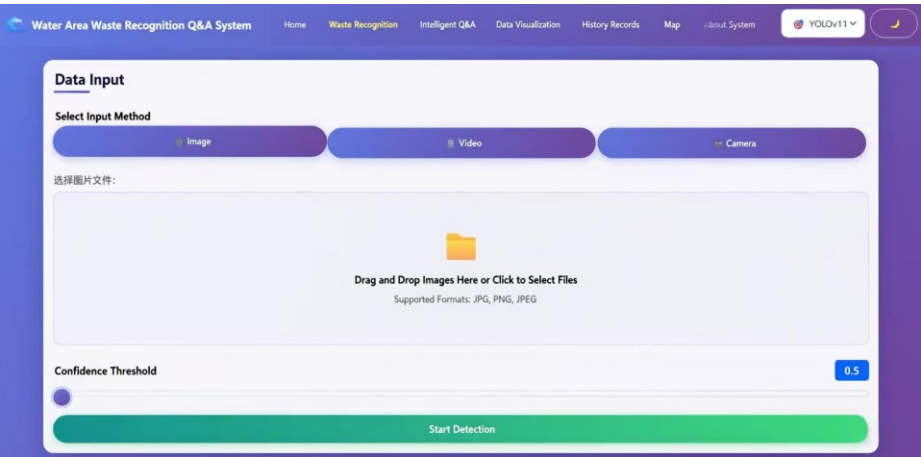


Figure 6. Recognition interface with model selection, confidence-threshold slider and detection trigger.

**Data input and pre-processing module**

The system supports drag-and-drop upload of Joint Photographic Experts Group (JPG) or Moving Picture Experts Group-4 Part 14 (MP4) files or capture of real-time images from the device camera.

Pre-processing crops borders and equalizes dim images. As shown in Figure 7, the historical detection records module lists each record’s time, location, model type and statistics in a table. It supports search, batch deletion and report export.

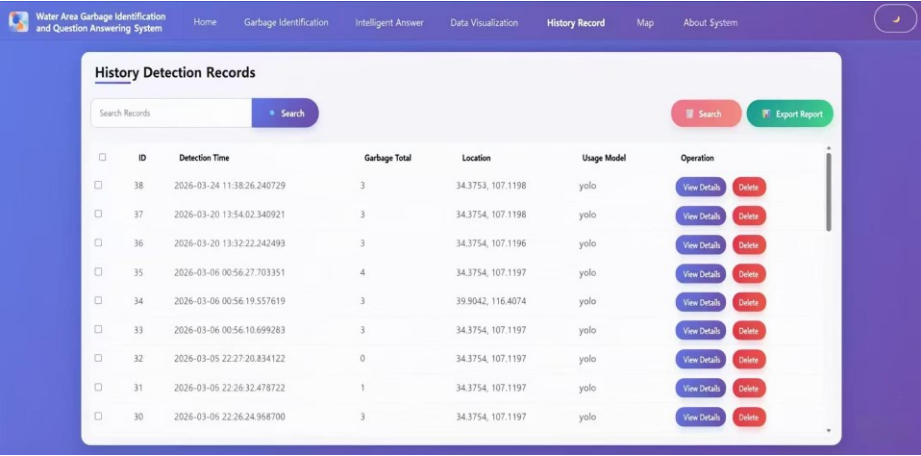


Figure 7. Historical detection records management module listing time, location, model and statistics.

**Garbage recognition and statistics module**

YOLOv11 achieves an average recognition accuracy of 78.30% over 11 garbage categories, and Faster R-CNN reaches 82.10%. The results are annotated with bounding boxes, and the right-side panel lists the quantity and

confidence of each category. As shown in Figure 8, every detected object is annotated with a bounding box, while a right-side panel lists the quantity and confidence of each garbage category in a structured list [8].

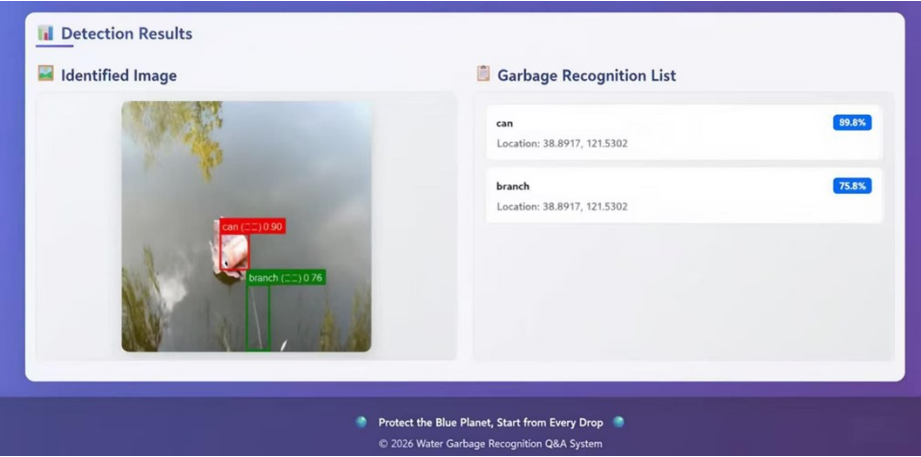


Figure 8. Garbage recognition results shown with annotated boxes and a per-category confidence list.

### Data analysis and visualization module

According to historical records, the system generates heat zones on a map, where deeper red indicates higher garbage density. It also provides temporal trend analysis and model comparison for decision support.

As shown in Figure 9, a clear bar chart summarizes the number of images and annotations for each of the eleven distinct garbage categories. This detailed visual provides a solid statistical basis for the subsequent analysis and evaluation.

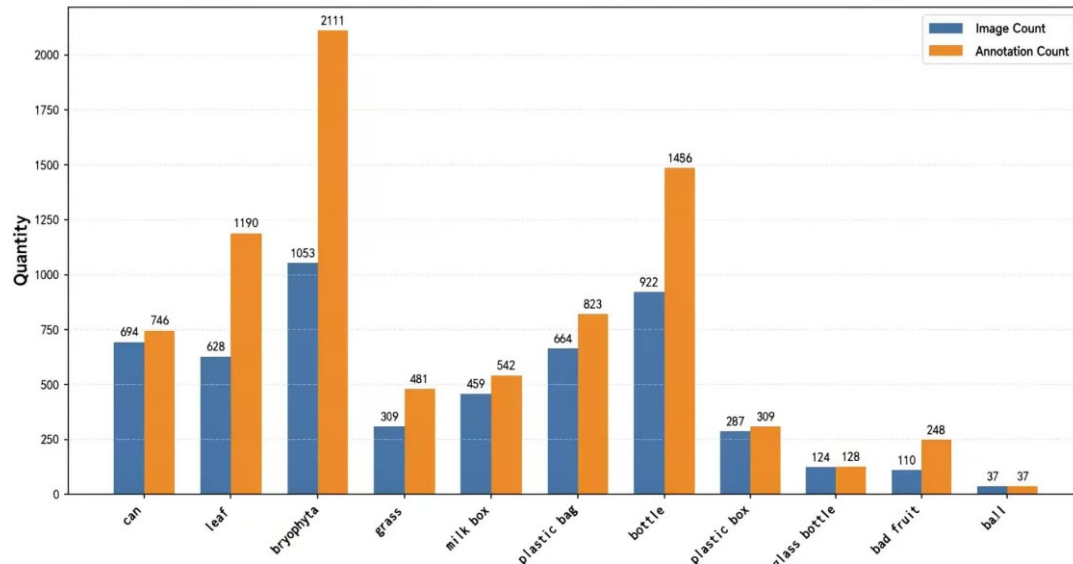


Figure 9. Sample distribution across the eleven garbage categories (number of images and annotations).

### Intelligent question-answering system

This module delivers intelligent explanations of water-area garbage, environmental-protection issues and garbage recycling, responding to user questions in natural language through the knowledge graph and large model.

### Innovation analysis

#### Innovations

The main innovations are best understood as threefold and mutually reinforcing. The first is a dual-model adaptive architecture that lets users switch between YOLOv11 for real-time patrol and Faster R-CNN for fine-grained analysis according to the scenario. This balances efficiency and accuracy. The second is question-answering interaction, which pioneers the “recognition + answering” mode. It lowers the usage barrier for non-professionals and enhances popular-science value. The third is spatio-temporal correlation analysis, which binds garbage data with time and Global Positioning System (GPS) coordinates. This effectively reveals pollution-diffusion patterns and strongly and reliably supports traceability [9].

#### Advancement and uniqueness

Compared with existing products, the system fuses

cutting-edge Computer Vision (CV) algorithms (YOLOv11), Natural Language Processing (NLP) technology (Large Language Model, LLM, question-answering) and geographic information systems (Leaflet). This resolves the pain points of “detecting without explaining” and “showing pictures without numbers”. It supports Docker deployment and adapts to hardware ranging from cloud servers to Raspberry Pi, offering strong flexibility.

#### Competitive advantage

The competitive advantages are three. First, a technical barrier is formed by three core capabilities: dual-model collaboration, real-time question answering, and full-process visualization. These are hard to replicate. Second, a cost advantage: The open-source technology stack lowers development expenses, and the hardware requirement is low; one Graphics Processing Unit (GPU) suffices. Third, market fit. It precisely addresses environmental agencies’ needs for “fewer people, higher efficiency”. It can also be extended to river-chief assessment and scenic-area cleaning management.

#### Application prospects and social value

##### Application prospects

In the short term, the system can be applied to urban river-channel patrol, replacing 30.00% of manual labor,

and to reservoir floating-object monitoring. In the medium term, it can extend to marine garbage census and evidence collection for illegal discharge. In the long term, combined with drones and 5th Generation Mobile Communication Technology (5G) networks, it can build a “sky-space-ground” integrated monitoring network for all-domain waters.

### Social value

The social value is threefold. First, governance efficiency is improved because recognition is about 50 times faster than manual work and runs 7×24 hours uninterruptedly. Second, environmental knowledge is popularized because the question-answering module reaches about one million users per year, strengthening public responsibility. Third, scientific decision-making is supported because data-driven pollution warning and resource allocation reduce pollution-control costs.

### Existing problems and future work

#### Existing problems

The remaining problems are threefold. The first is complex interference, where strongly reflective water surfaces and floating vegetation easily cause false alarms. Under low nighttime illumination, the missed-detection rate rises. The second is limited model generalization, where recognition of unseen new garbage is insufficient, for example, new takeout-box packaging. The third is an edge-side bottleneck, where inference on devices such as Raspberry Pi reaches only 2.00 to 3.00 fps. This can hardly meet real-time requirements [10].

#### Future improvement directions

Future work will focus on four complementary and practical directions. The first is data augmentation that synthesizes fog, wave and other disturbance samples to improve robustness. The second is incremental learning that supports online model updates to adapt to new garbage types. The third is model distillation that compresses Faster R-CNN knowledge into lightweight networks to enhance edge-side performance. The fourth is multi-modal fusion that combines infrared cameras to solve night-time detection under low-light conditions.

### Conclusion

This project successfully builds a complete water-area garbage recognition and question-answering system.

<https://www.wonford.com/>

Through the deep fusion of computer vision and natural language processing, it realizes a “perception - cognition - interaction” closed loop. In multi-scenario tests, the system remains stable, achieving a recognition accuracy above 75.00% and a question-answering accuracy of 88.00%, demonstrating potential for large-scale promotion. Future work will continue to optimize model adaptability and broaden application boundaries, contributing technology to the protection of clean waters.

### Funding

This work was not supported by any fund.

### Acknowledgements

The authors would like to show sincere thanks to those techniques who have contributed to this research.

### Conflicts of Interest

The authors declare no conflict of interest.

### References

- [1] Chen, L., Zhu, J. (2024) Water surface garbage detection based on lightweight YOLOv5. *Scientific Reports*, 14(1), 6133.
- [2] Jiang, P., Ergu, D., Liu, F., Cai, Y., Ma, B. (2022) A review of Yolo algorithm developments. *Procedia Computer Science*, 199, 1066-1073.
- [3] Yi, Z., Yao, D., Li, G., Ai, J., Xie, W. (2022) Detection and localization for lake floating objects based on CA-faster R-CNN. *Multimedia Tools and Applications*, 81(12), 17263-17281.
- [4] Ameen, H., Soilán, M. (2026) A review of large language models in geomatics: integrating multimodal data, addressing challenges, and exploring synergies. *Spatial Information Research*, 34(2), 16.
- [5] Hao, C. Y., Chen, Y. C., Chen, T. T., Lai, T. H., Chou, T. Y., Ning, F. S., Chen, M. H. (2024) Synthetic data-driven real-time detection transformer object detection in raining weather conditions. *Applied Sciences*, 14(11), 4910.
- [6] Yang, Y., Liu, X., Tu, X., Lu, Y., Wang, Y. (2025) Automating the construction of environmental policy knowledge graph with large language models. *Sustainability*, 17(22), 10282.
- [7] Purkayastha, K. D., Mishra, R. K., Shil, A., Pradhan, S. N. (2021) IoT based design of air quality

- monitoring system web server for android platform. *Wireless Personal Communications*, 118(4), 2921-2940.
- [8] Edozie, E., Shuaibu, A. N., John, U. K., Sadiq, B. O. (2025) Comprehensive review of recent developments in visual object detection based on deep learning. *Artificial Intelligence Review*, 58(9), 277.
- [9] Mohan, S., Kumar, B., Nejadhashemi, A. P. (2025) Integration of machine learning and remote sensing for water quality monitoring and prediction: a review. *Sustainability*, 17(3), 998.
- [10] Gou, J., Yu, B., Maybank, S. J., Tao, D. (2021) Knowledge distillation: a survey. *International Journal of Computer Vision*, 129(6), 1789-1819.

## Appendix

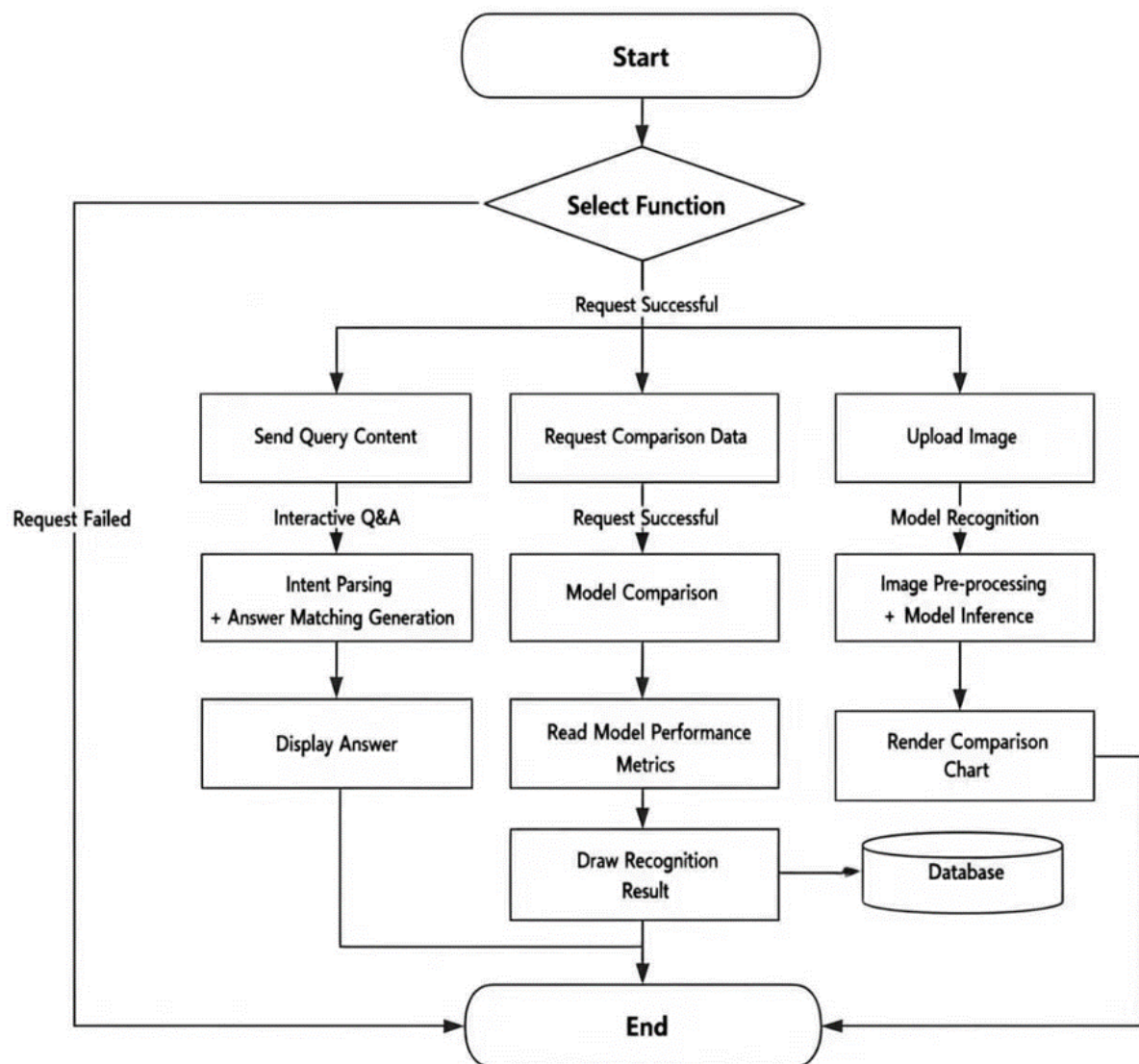


Figure A1. Core business workflow covering garbage recognition, intelligent question-answering and data comparison.