Earth, Sea and Sky XI:
International Joint Graduate Program Workshop in Earth and Environmental Sciences
hosted by The International Joint Graduate Program in Earth and Environmental Sciences (GP-EES), Tohoku University
June 1 – June 4, 2025

Semantic Plankton: Multimodal LLMs and RAG System for Automated Marine Plankton Analysis

Jaronchai Dilokkalayakul[1], Akane Kitamura[2], Takeshi Obayashi[1,2]

[1] Graduate School of Information Sciences, Tohoku University
[2] Advanced Institute for Marine Ecosystem Change (WPI-AIMEC), Tohoku University

Plankton is foundational to marine ecosystems, serving as sensitive indicators of ocean health and environmental change. However, traditional identification methods are labor-intensive and difficult to scale. While some AI-powered image classification tools exist, they often rely on static image datasets and lack the ability to incorporate contextual information. In this study, we present a system for automated plankton analysis that leverages semantic understanding through multimodal data. Here, "semantic" refers to the system's ability to integrate visual features with contextual and behavioral metadate, which leverages multimodal large language models (LLMs) and retrieval-augmented generation (RAG).

The system integrates the *PlanktoScope*, a Raspberry Pi-based imaging device, to capture high-resolution video of marine samples. These are processed through a multi-stage image pipeline involving contrast enhancement, background subtraction, thresholding, and morphological operations. Individual plankton images are segmented, tracked across frames, and annotated with spatial and size metadata.

For classification, we employ a multimodal LLM capable of interpreting both visual input and structured metadata to generate detailed, taxonomically relevant output in JSON format. To provide contextual grounding, a RAG mechanism retrieves semantically similar plankton instances and metadata from a custom-built local vector database of curated images, improving classification reliability. This also allows the system to account for behavioral traits, such as movement patterns and environmental influences. A diagram of the LLM and RAG workflow is shown in Figure 1.

Our preliminary results suggest that the multimodal LLM coupled with RAG approach yields promising performance in classifying plankton. Its ability to incorporate contextual and environmental information, along with structured metadata, shows potential for more nuanced and accurate classification. Furthermore, the system is designed to be modular and scalable, with real-time or near-real-time capabilities, making it well-suited for future integration into large-scale environmental monitoring networks.

This project shows the potential of combining modern technology of generative AI, and RAG techniques for marine biodiversity research and aims to reduce manual taxonomic labor through AI-augmented observation. Future applications include functional annotation of marine microbiomes, automated monitoring in fish farms, and broader ecological trend analysis across time/geography.

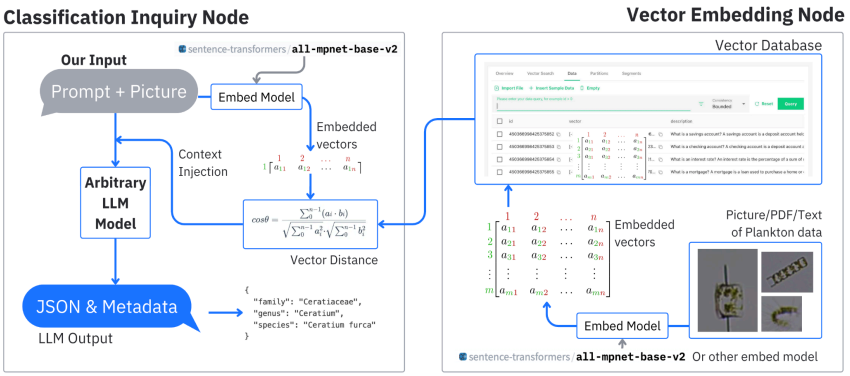

Figure 1. Overview of the LLM and RAG workflow