

Abstract Draft for EAFES2025

Context-Aware Marine Plankton Classification with Multimodal Large Language Model and Retrieval-Augmented Generation Reasoning

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

* : Corresponding author

Marine plankton are vital to ocean ecosystems and serve as indicators of biodiversity and environmental change. Scalable monitoring of plankton communities is essential for understanding ecological dynamics and the effects of climate change. However, traditional identification through manual microscopy is time-consuming, expertise-dependent, and not scalable.

Convolutional Neural Networks (CNNs) have been used to automate plankton image classification, but they require large labeled datasets. They also struggle to generalize across variations in sampling and imaging conditions, such as lighting, zoom level, or filter pore size, as well as environmental metadata that human experts often consider in classification.

To address these limitations, we propose a framework that combines Multimodal Large Language Models (LLMs) with Retrieval-Augmented Generation (RAG). LLMs enable reasoning over both visual and textual inputs, allowing context-aware classification that integrates domain-specific knowledge, including metadata and imaging conditions. While pre-trained LLMs reduce the need for large labeled datasets, RAG further enhances the process by incorporating domain-specific knowledge through the retrieval of similar examples from a curated plankton database.

Our system processes video microscopy data from PlanktoScope by detecting individual plankton, pairing each instance with metadata-informed prompts, and retrieving comparable samples for inference. The LLM then outputs structured taxonomic predictions in JSON format, including genus and family.

Designed with a modular architecture, the system supports scalable deployment and can integrate with real-time biodiversity monitoring platforms. Preliminary results demonstrate promising accuracy. Beyond marine applications, the framework is generalizable to domains such as microbiome analysis and automated species annotation. By reducing manual workload and enhancing interpretability, our approach advances biological classification and contributes to more responsive environmental monitoring.

Keywords: Planktons ; Marine Plankton ; Large Language Models (LLM) ; Retrieval Augmented Generation (RAG) ; Vector Embedding ; Semantic Search ; Marine Biodiversity Monitoring ; Image Based Classification ; PlanktoScope ; Automated Taxonomic Classification