Semantic Plankton: Multimodal LLMs and RAG for Automated Ocean Microorganism Analysis

AUTOMATIC IMAGE ANALYSIS WITH PLANKTOSCOPE IN THE PLANDYO PROJECT

Jaronchai Dilokkalayakul [1] , Akane Kitamura [2] , Takeshi Obayashi [1,2] ([1]Graduate School of Information Sciences, Tohoku University | [2]WPI-AIMEC)

Summary

Traditional plankton identification methods are labor-intensive and hard to scale, posing challenges for monitoring marine ecosystems and detecting environmental changes. While CNNs automate classification, they depend on large labeled datasets and lack contextual reasoning. To address this, we propose a framework that combines large language models (LLMs) with retrieval-augmented generation (RAG) to classify plankton using both image and contextual metadata. By retrieving semantically similar examples from a curated vector database and integrating them into the Large Language Model's input, our system enables context-aware, scalable classification with reduced reliance on labeled data.

Process Pipeline

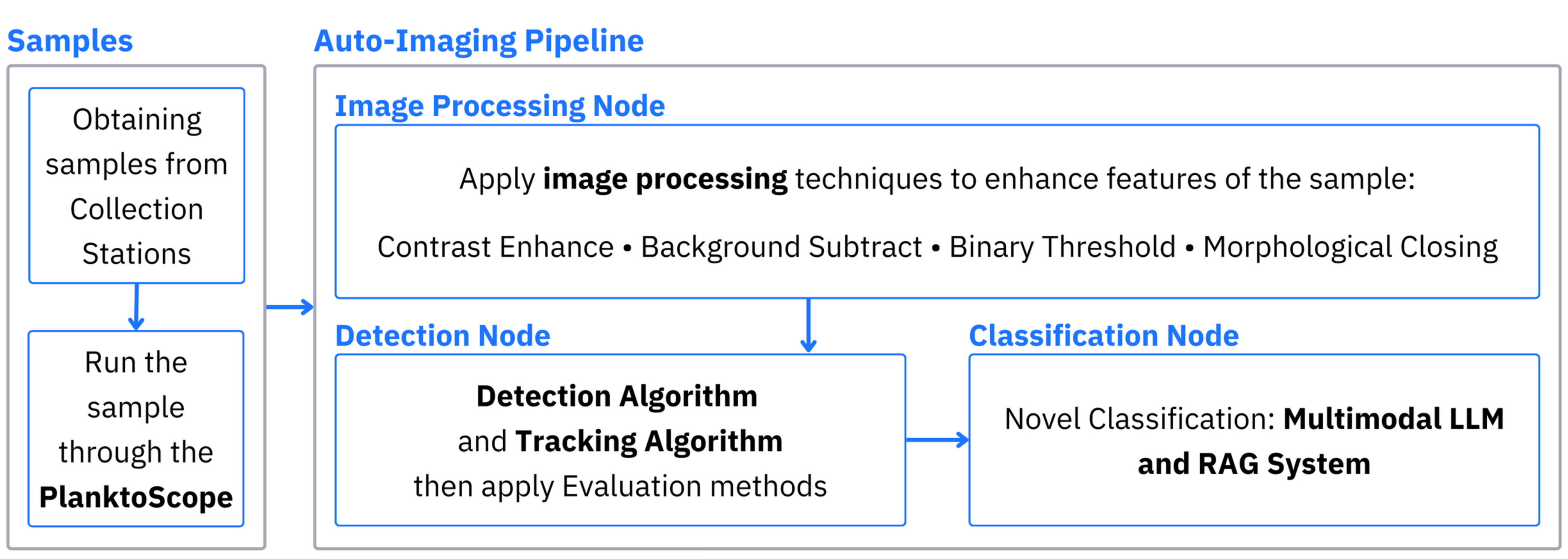
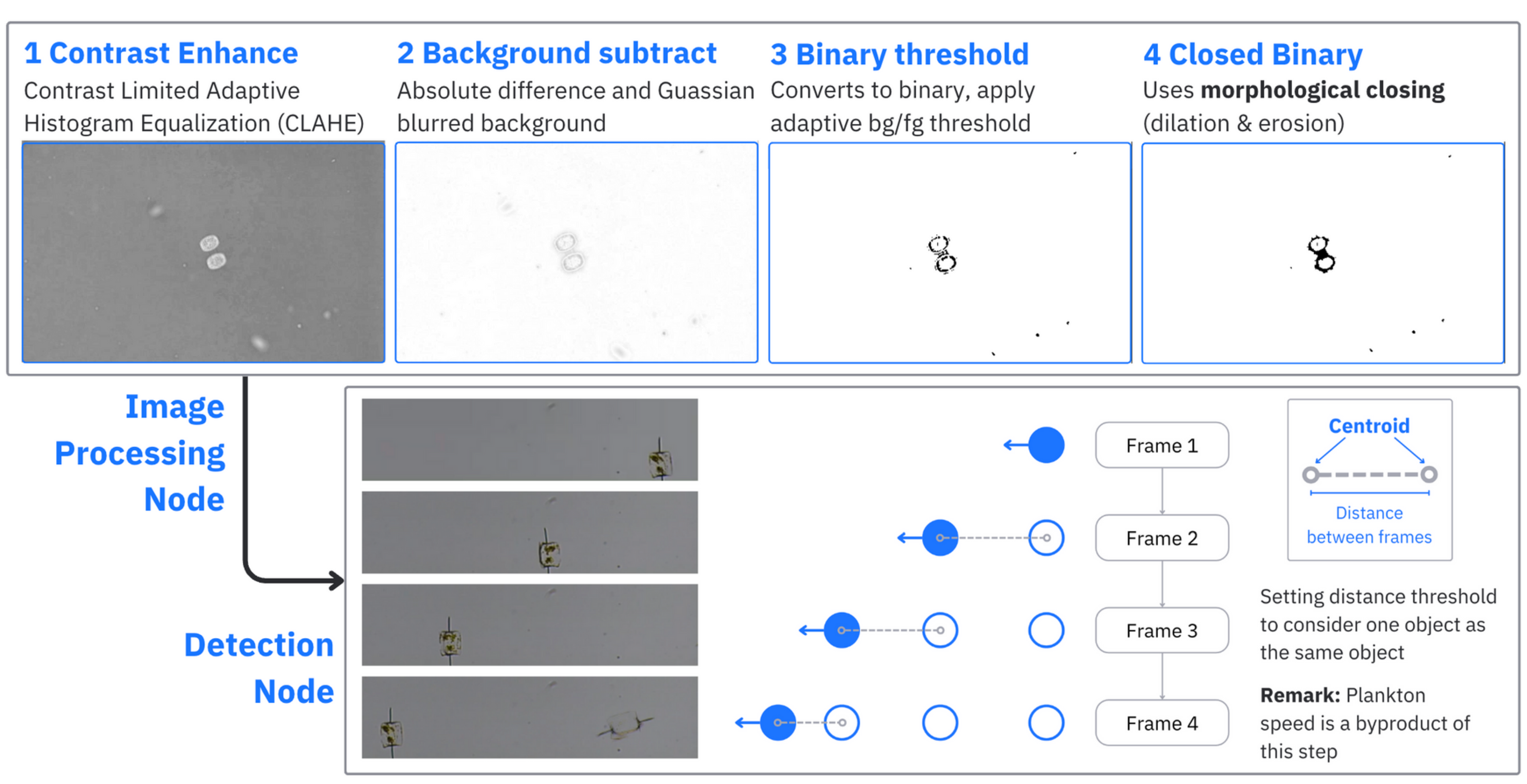
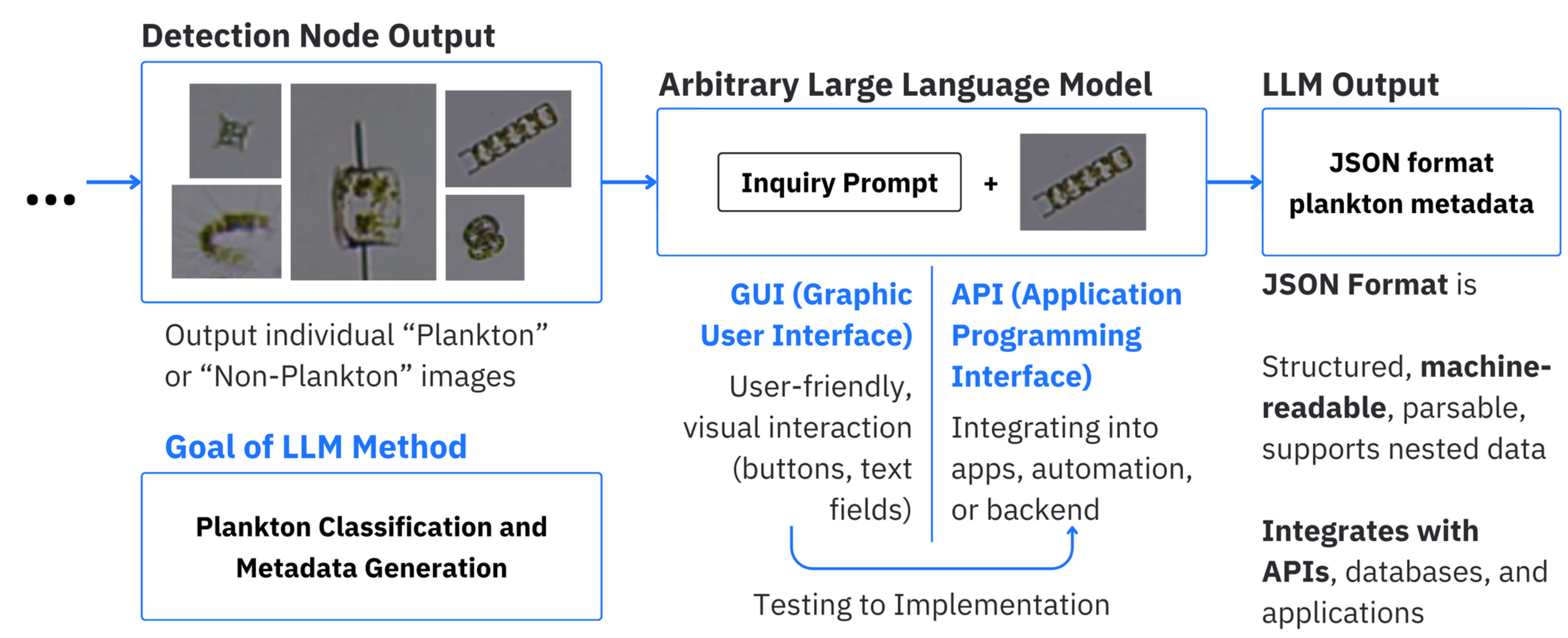


Image Processing



Key Concept of LLM Inquiry



Advantage of LLM Implementation

Feature	Traditional Image Classification	Large Language Model + RAG System
Input Data	Static images	Continuous video frames
Output Format	Static segmentation	Context-aware, richer taxonomic outputs
Motion Analysis	✕ Not possible	Tracks movement trajectories
Environmental Factors	✕ Not considered	Includes temperature, pH, angle
Behavioral Insights	✕ Limited to morphology	Behavior changes based on factors

Example Output Metadata Results

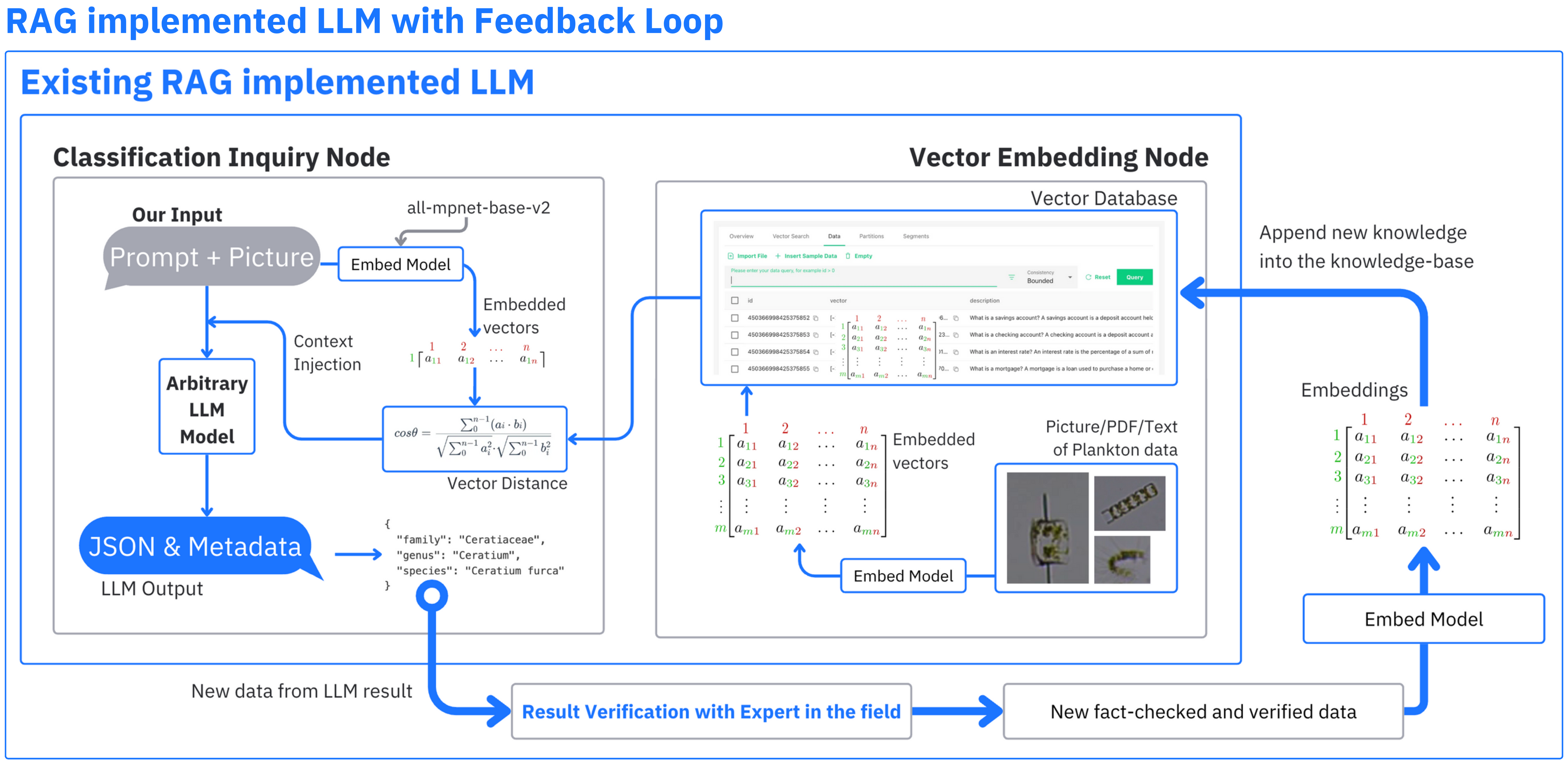
{
"family": "Chaetocerotaceae",
"genus": "Chaetoceros"
}

{
"family": "Ditylaceae",
"genus": "Ditylum"
}

{
"family": "Ceratiaceae",
"genus": "Ceratium"
}

Large Language Model and Retrieval-Augmented Generation Framework

RAG implemented LLM with Feedback Loop



A prompt and an image of a plankton specimen are provided as the input query. Then the image and textual context are encoded into vector embeddings. The embedding is used to search a vector database containing embedded plankton documents, this retrieves semantically similar examples. The retrieved context is then injected into an LLM, which uses both the input and retrieved information to generate a structured JSON taxonomic metadata output. New outputs are verified by experts to confirm the result. Validated examples are embedded and added back to the vector database, as a continuous improvement to the system's contextual grounding.

Contact Information

Presenter: **Jaronchai Dilokkalayakul**
Email: dilokkalayakul.jaronchai.p8@dc.tohoku.ac.jp
Personal Website: https://jaronchai.com