Contribution ID: **39**   Type: **Long Talk**

iMagine: Best practices for suppliers of image collections and analysis tools in aquatic sciences

Thursday, 3 October 2024 09:30 (20 minutes)

The iMagine platform utilizes AI-driven tools to enhance the processing and analysis of imaging data in marine and freshwater research, supporting the study of crucial processes for ocean, sea, coastal, and inland water health. Leveraging the European Open Science Cloud (EOSC), the project provides a framework for developing, training, and deploying AI models. To effectively achieve the objectives of the project, about twelve use cases in different areas of aquatic science are collaborating with the providers of the iMagine AI platform. This collaboration has yielded valuable insights and practical knowledge. Thoroughly revising the existing solutions from data acquisition and preprocessing to the final stage, provides a trained model as a service to the users.

Within the framework of iMagine, we outline various tools, techniques, and methodologies appropriate for aquatic science image processing and analysis. In this work, we delve into the best AI-based solutions for image processing, drawing on the extensive experience and knowledge we have gained over the course of the iMagine project. Clear guidelines for annotating images, coupled with comprehensive training and accessible tools, ensure consistency and accuracy in labeling. Thus, We verify annotation tools such as BIIGLE, Roboflow, LabelStudio, CVAT, and LabelBox based on the different features, along with real-time video streaming tools.

Preprocessing techniques and quality control measures are discussed to enhance data quality in aquatic datasets, aiming to identify and address issues such as blurriness, glare, or artifacts. Preparation of training datasets and their publishing in a data repository with the relevant metadata is assessed. Following this, an overview of deep learning models, including convolutional neural networks, and their applications in classification, object detection, localization, and segmentation methods is provided.

Performance metrics and evaluation methods, along with experiment tracking tools such as Tensorboard, MLflow, Weight and Biases, and Data Version Control are discussed for the purpose of reproducibility and transparency. Ground truth data is utilized to validate and calibrate image analysis algorithms, ensuring accuracy and reliability. Furthermore, AI model drift tools, data biases, and fairness considerations in aquatic science models are discussed, concluding with case studies, discussions on challenges and limitations in AI applications for aquatic sciences.

By embracing these best practices, providers of image collections and image analysis applications in aquatic sciences can enhance data quality, promote reproducibility, and facilitate scientific progress in this field. A collaboration of research infrastructures and IT experts within the iMagine framework results in the development of best practices for delivering image processing services. The project establishes common solutions in data management, quality control, performance, integration, and FAIRness across research infrastructures, thereby promoting harmonization and providing input for best practice guidelines.

Finally, iMagine shares its developments with other leading projects such as AI4EOSC and Blue-Cloud 2026 to achieve optimal synergy and wider uptake of the iMagine platform and best practices by the larger aquatic and AI research communities.

Topic

Needs and solutions in scientific computing: Artificial Intelligence

Primary authors: Dr AZMI, Elnaz (Karlsruhe Institute of Technology (KIT)); Dr ALIBABAEI, Khadijeh (Karlsruhe Institute of Technology (KIT)); Dr KOZLOV, Valentin (Karlsruhe Institute of Technology (KIT)); Dr LOPEZ GARCIA, Alvaro (CSIC); Dr SCHAAP, Dick (Mariene Informatie Service MARIS BV); Dr SIPOS, Gergely (EGI.eu)

Presenter: Dr AZMI, Elnaz (Karlsruhe Institute of Technology (KIT))

Session Classification: Processing Research Data with Artificial Intelligence and Machine Learning