Internship - Artificial Intelligence for Environment: Generative deep learning models for data completion in Oceanography

Topic : Artificial Intelligence and Environment

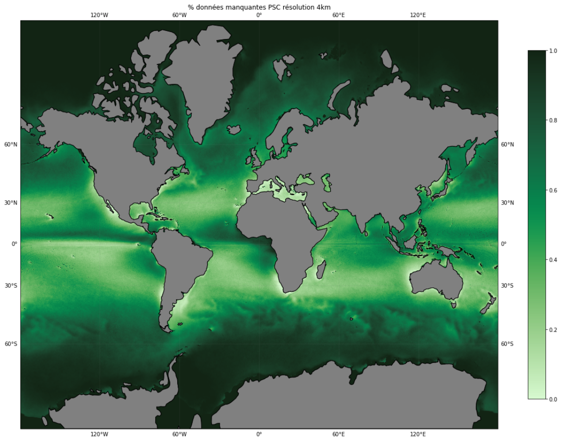
Contact: patrick.gallinari@sorbonne-universite.fr, sylvie.thiria@locean.ipsl.fr, luther.ollier@locean.ipsl.fr

Location: ISIR (Institut des Systrèmes Intelligents et de Robotique) and LOCEAN (Laboratoire d'Océanographie et du Climat:) 4 Place Jussieu, 75005 Paris

Profile : Engineering school or master degree in applied maths, computer science or environment science.

When: starting 2024 Spring for 6 months

1. Context: Phytoplankton, the microscopic marine organisms, play a vital role in ocean ecosystems by producing a significant portion of the world's oxygen and serving as the base of the marine food web. Data regarding sea surface phytoplankton ratios are primarily derived from satellite observations and in situ samples [1]. However, despite technological advancements, there are still substantial missing points within these datasets.



Several studies have been conducted to address this issue [2,3], but there remains untapped potential in leveraging generative deep learning models, such as variational auto-encoders or diffusion models, to enhance our understanding of phytoplankton distribution and dynamics.

2. Proposed Study: The proposed internship project will center on exploring the application of generative models for filling gaps in the existing sea surface phytoplankton ratio database. Generative deep learning models offer promising capabilities in filling missing or incomplete data while preserving the underlying structure and distributions. This study aims to harness the potential of selected generative models (e.g. VAEs [4] and diffusion models [5]) to effectively impute and reconstruct missing phytoplankton ratio information from incomplete satellite observations and in situ samples. The primary challenge is to work in the absence of a ground truth database and discover novel methods to validate the results

3. Data:

- PSC : 8640*4320*1068 (lon, lat, time) → 60 % missing points

- CHL : 8640*4320*1068 (lon, lat, time) → 10 % missing points

Intern Responsibilities:

- Collaborate with our research team to understand existing phytoplankton datasets and their limitations.
- Explore and implement generative models suitable for the specific characteristics of phytoplankton ratio data.
- Develop and fine-tune algorithms for inpainting missing data points in the phytoplankton dataset.
- Evaluate the performance of the generative models through quantitative and qualitative analysis, comparing inpainted data against known ground truth.
- Document methodologies, findings, and recommendations in a clear and concise manner for internal reference and potential publication.

Qualifications:

- Pursuing a degree (Master's) in Computer Science, Applied Maths, Data Science, Environmental Science, or a related field.
- Proficiency in recent machine learning techniques.
- Strong programming skills in languages such as Python.
- Excellent problem-solving skills and a passion for leveraging technology to address environmental challenges.

Benefits:

- Hands-on experience in applying cutting-edge machine learning techniques to address real-world environmental data challenges.
- Mentorship from experienced researchers in both machine learning and environmental science domains.
- Opportunity to contribute to impactful research aimed at enhancing our understanding of marine ecosystems.
- Potential for co-authorship on publications resulting from the internship project.

Duration & Location:

- The internship is a 6 months commitment, onsite at ISIR or possibly at LOCEAN (4 pl Jussieu 75005 Paris).

Application Process:

- Interested candidates are invited to submit their resume, a cover letter highlighting their relevant experience and interest in the project, and any supporting materials (e.g., portfolio, GitHub repository) to patrick.gallinari@sorbonne-universite.fr and sylvie.thiria@locean.ipsl.fr.

We look forward to welcoming a passionate and driven intern to join our team in exploring the innovative application of variational auto-encoders to advance our understanding of sea surface phytoplankton ratios and their implications for ocean health.

References :

[1] Roy El Hourany. Télédétection du phytoplancton par méthode neuronale : du global au régional, de la composition pigmentaire aux biorégions. Biodiversité et Ecologie. Sorbonne Université, 2019. Français. NNT : 2019SORUS095. tel-02562426v2
[2] Matthew Ehrler, Neil Ernst VConstruct: Filling Gaps in Chl-a Data Using a Variational Autoencoder
[3] Joana Roussillon, Ronan Fablet, Thomas Gorgues, Lucas Drumetz, Jean Littaye, Elodie Martinez A Multi-Mode Convolutional Neural Network to reconstruct satellite-derived chlorophyll-a time series in the global ocean from physical drivers
[4] Mark Collier, Alfredo Nazabal, Christopher K.I. Williams, VAEs in the Presence of Missing Data 2021
[5] A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte and L. Van Gool, "RePaint: Inpainting using Denoising Diffusion Probabilistic Models," 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), , 2022, pp. 11451-11461