Environmental Computational Science
and Earth Observation Laboratory

Master thesis subject

Species distribution modelling with multi-modal deep learning

Context

Describing and understanding the geographic distribution and environmental suitability of species is a central question in ecology and biogeography. It has become increasingly meaningful in the face of anthropogenic pressure on biodiversity and has occasioned the development of **species distribution models (SDMs)**. Such models relate species occurrence data with environmental variables and are used to understand and predict species' distributions across landscapes (Beery et al., 2021).

At the same time, **machine learning** and especially **deep learning** methods have become increasingly popular in many fields, including ecology. However, their application to SDM has not yet been widely adopted. This is partly due to the fact that the datasets of species observations used in SDMs are small and suffer from high selection biases, making the use of naive deep learning approaches difficult. The GeoLifeCLEF 2023 competition aims to address these issues by proposing an extensive multi-modal dataset of 5M of species observations together with satellite imagery, environmental rasters, and climatic and multi-spectral time series.

Project

We aim to investigate the effectiveness of multi-modal deep learning methods for improving the accuracy and robustness of SDMs using the GeoLifeCLEF 2023 dataset. Specifically, we plan to explore the use of convolutional neural networks (CNNs) or transformers for processing satellite imagery and environmental rasters, and recurrent neural networks (RNNs) or transformers for processing time series data. The main challenges of this project are to deal with the large size and the different modalities of the dataset, and to handle the training of the different large neural networks.

ECEO

EPFL ENAC IIE
Devis Tuia, Prof.
Rue de l'Industrie 17
Case Postale 440
CH – 1951 Sion

Téléphone : +41 21 69 382 83 (secr).
E-mail : devis.tuia@epfl.ch

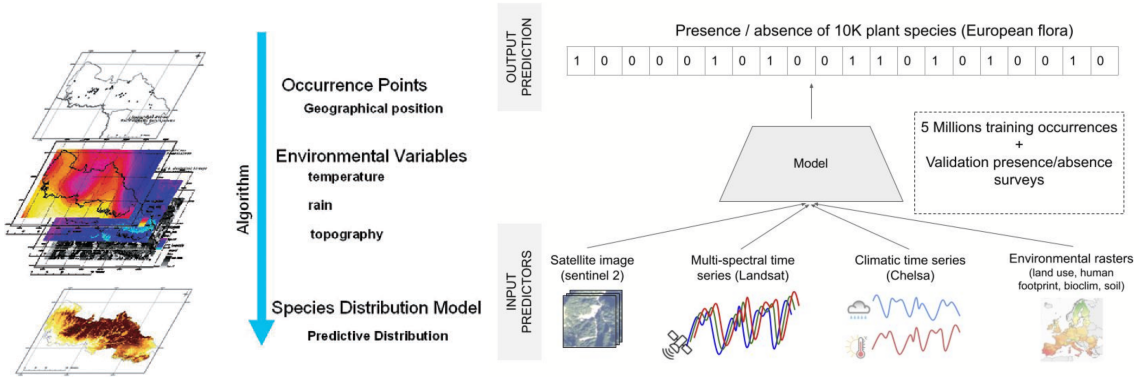

Figure: (Left) An example of species distribution modelling. Presence only data is used to sample environmental covariates of where a species might occur. A machine learning model then extrapolates out the correlative relationship between presence and environment to generate a habitat suitability map. (Fook et al, 2009). (Right) Setting of the GeoLifeCLEF 2023 competition.

Requirements

- Experience in deep learning
- Proficiency in Python and relevant libraries such as Scikit-learn and Pytorch
- Familiarity with ecology and SDMs is a plus
- Strong willingness to learn and ability to work independently

Literature

- Sara Beery, Elijah Cole, Joseph Parker, Pietro Perona, and Kevin Winner. 2021. Species Distribution Modeling for Machine Learning Practitioners: A Review. In ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS '21). Association for Computing Machinery, New York, NY, USA, 329–348. https://doi.org/10.1145/3460112.3471966

Contact

Prof. Devis Tuia, devis.tuia@epfl.ch