UNIVERSITY OF CAMBRIDGE

Towards Scalable Deep Species Distribution Modelling Using Global Remote Sensing

Emily F. Morris[1], Anil Madhavapeddy[1], Sadiq Jaffer[1], David Coomes[2]

[1] Department of Computer Science, University of Cambridge [2] Department of Plant Sciences, University of Cambridge

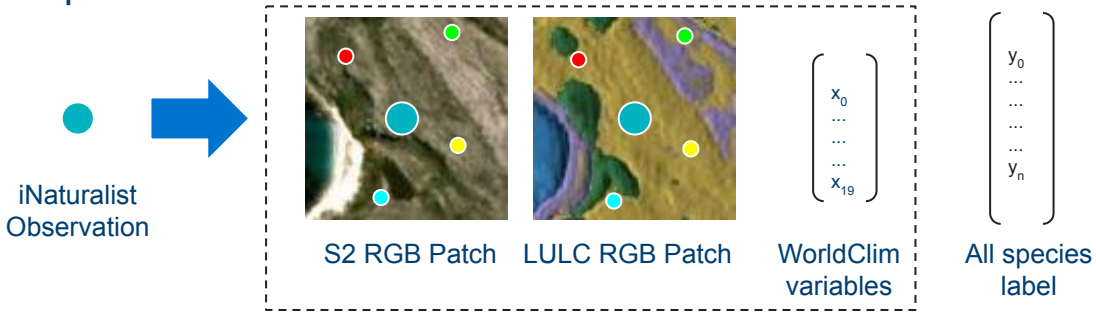
Department of Computer Science

Introduction

- Destruction of natural habitats and anthropogenic climate change are threatening biodiversity globally.
- Addressing this loss necessitates enhanced monitoring techniques to assess the impact of environmental shifts and to guide policy-making efforts.
- **Aim: to provide an improved, scalable method for species distribution modelling by incorporating global remote sensing data and using this with CNN-based architecture**
- Investigate performance using Protea species in the Cape Floristic Region of South Africa

Methods & Models

- Data:
  - Adapt dataset pipeline [1] that creates patches centred on each observation. Each patch is labelled with all the species occurring in the patch.
  - Data point example:



- Models:
  - *Deepbiosphere* [1], a multimodal version of a T-ResNet CNN model
  - Maxent, classic statistical model used for species distribution modelling

Training splits

- Spatial autocorrelation between evaluation and training sets can cause overly optimistic results.
- Create representative train/validation splits while addressing spatial autocorrelation

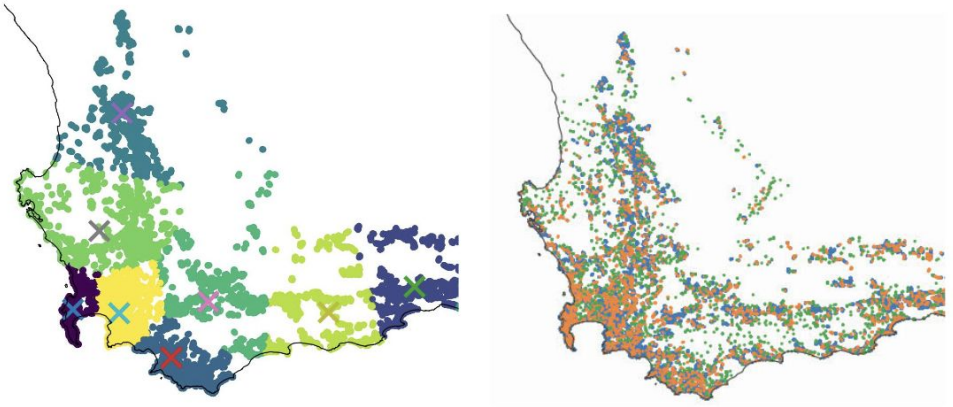


Figure 3: Example clusters for part of the Cape Floristic Region used in creating the validation sets (*left*) and separation of points into train (blue), validation (orange) and test (blue) sets for the region (*right*).

Results & Conclusions

- The CNN models perform better on average for AUC_{ROC}:
  - Image + environmental mean AUC_{ROC}: **0.8849 ± 0.046**
  - Image + LULC + environmental mean AUC_{ROC}: **0.8822 ± 0.048**
  - Maxent mean AUC_{ROC}: **0.8550 ± 0.059**
- Overall we show that we can train a model to make predictions based on global remote sensing data that has fine-grained spatial and temporal scales.
- Results need to be further investigated using a field study and ground truth data.

[1] Gillespie, L., Ruffley, M. and Expósito-Alonso, M., 2021. An Image is Worth a Thousand Species: Scaling high-resolution plant biodiversity prediction to biome-level using citizen science data and remote sensing imagery. *Biodiversity Information Science and Standards*, *5*, p.e74052.