Predicting Species Occurrence Patterns from Partial Observations

Hager Radi Abdelwahed[1], Mélisande Teng[1,2], David Rolnick[1,3]

[1]Mila Quebec AI Institute [2]Université de Montréal [3]McGill University

Motivation

- Climate change is a major threat to **species biodiversity** which is essential to maintain healthy ecosystems.
- Goal is to predict species occurrence patterns given:
  (a) satellite imagery
  (b) known information on the occurrence of other species.

Opportunity

ML + remote sensing + citizen science + available observations about other species

Task

Predict **jointly** species **encounter rates** from remote sensing data

Target at hotspot *h*:

$$y_s^h = \frac{\text{Number of complete checklists reporting } s \text{ at } h}{\text{Total number of complete checklists at } h}$$

Contributions

- **SatButterfly**: A dataset for predicting butterfly species encounter rates from remote sensing and environmental data.
  - A subset of the data are collocated with existing data from SatBird [Teng et. al 2023]

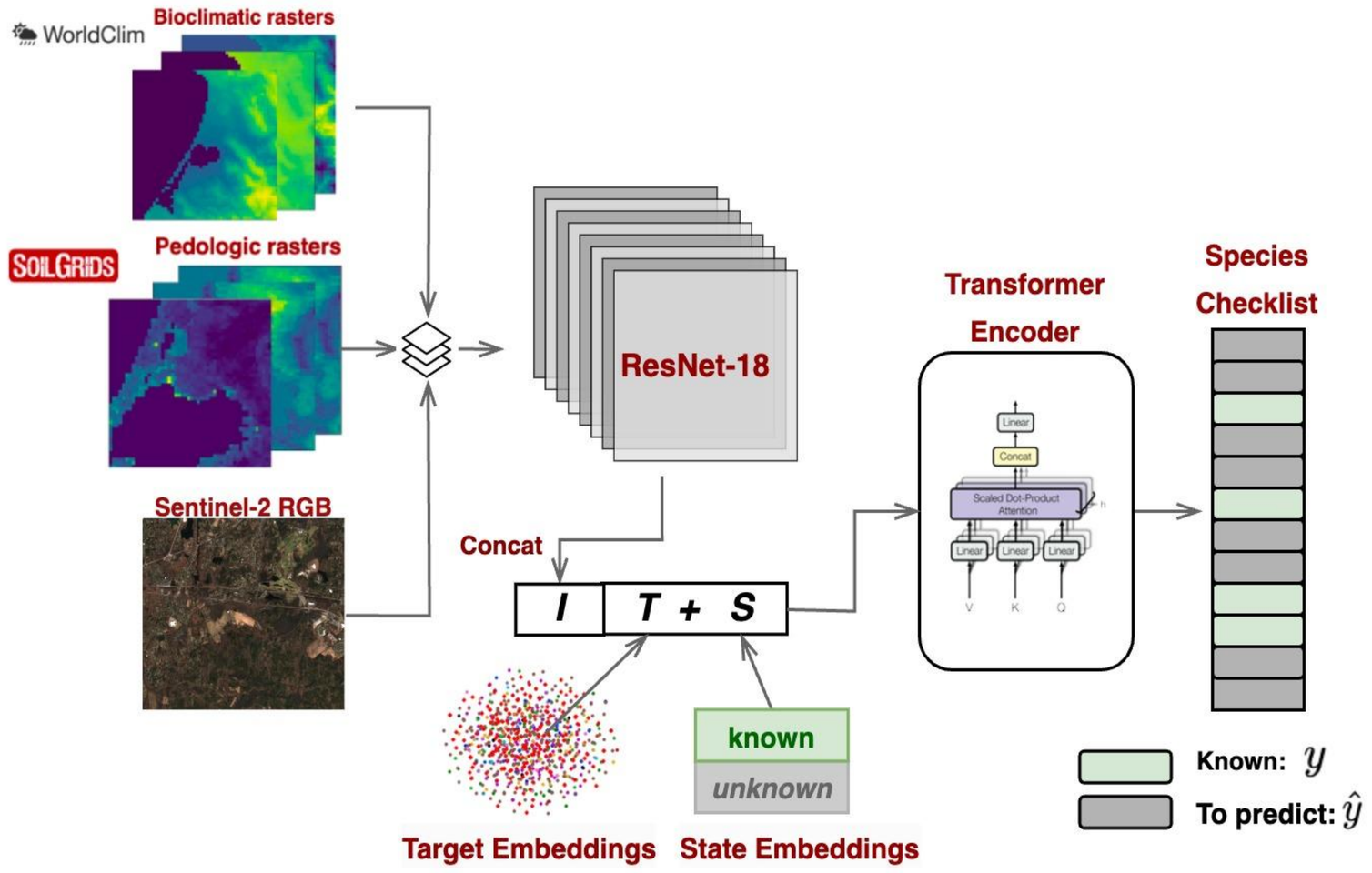
	SatButterfly-v1	SatButterfly-v2
train	5316	4677
validation	1147	1002
test	1145	1005

- **R-Tran**: A model to train and predict species encounter rates from partial information about other species.
  - R-Tran uses a transformer encoder architecture to model interactions between features of satellite imagery and targets.

References

- Feedback-prop: Convolutional neural network inference under partial evidence, CVPR, 2018.
- General multi-label image classification with transformers, CVPR 2021.
- Satbird: a dataset for bird species distribution modeling using remote sensing and citizen science data, NeurIPS 2023

Model



Results

- Feedback-prop is an inference method that utilizes partial information.

Table 1: Evaluation within a taxon (SatBird): *A* refers to the subset of songbird classes only, *B* refers to the subset of non-songbird classes.

Model	MAE[1^{-2}]	MSE[1^{-2}]	Top-10 %	Top-30 %	Top-k %
All bird species					
ResNet18	2.19	0.65	45.28 ± 0.61	64.62 ± 0.5	66.26 ± 0.43
R-Tran	2.13	0.67	44.5 ± 0.4	64.2 ± 0.3	66.1 ± 0.26
Evaluation on classes *A*: Songbirds					
ResNet18	3.25	1.02	53.75 ± 0.66	78.94 ± 0.42	69.92 ± 0.48
ResNet18-*A*	3.23	1.02	53.91 ± 0.23	79.03 ± 0.28	70.1 ± 0.29
Feedback-prop $(A\|B)$	2.97	0.90	56.54 ± 0.6	80.71 ± 0.35	71.72 ± 0.32
R-Tran $(A\|B)$	**2.46**	**0.72**	**59.98** ± 0.16	**82.3** ± 0.1	**73.45** ± 0.2
Evaluation on classes *B*: non-songbirds					
ResNet18	1.35	0.37	59.97 ± 0.41	86.14 ± 0.32	59.72 ± 0.35
ResNet18-*B*	1.32	0.36	60.04 ± 0.29	86.09 ± 0.19	59.61 ± 0.16
Feedback-prop $(B\|A)$	1.22	0.32	63.09 ± 0.67	87.68 ± 0.4	61.94 ± 0.48
R-Tran $(B\|A)$	**0.99**	**0.26**	**66.25** ± 0.08	**89.28** ± 0.07	**64.19** ± 0.11

Table 2: Evaluation across taxa (SatBird and SatButterfly): *A* refers to bird species, *B* refers to butterflies. Results are evaluated on the subset of locations with both bird and butterfly data available.

Model	MAE[1^{-2}]	MSE[1^{-2}]	Top-10 %	Top-30 %	Top-k %
Evaluation on classes *A*: Birds					
ResNet18	2.07	0.62	48.02 ± 0.2	65.9 ± 0.15	73.17 ± 0.15
Feedback-prop $(A\|B)$	2.06	0.62	48.11 ± 0.16	66.18 ± 0.19	73.24 ± 0.1
R-Tran $(A\|B)$	**2.04**	**0.6**	**50.34** ± 0.3	**67.24** ± 0.4	**73.85** ± 0.18
Evaluation on classes *B*: Butterflies					
ResNet18	3.81	1.55	52.35 ± 0.43	83.74 ± 0.43	35.69 ± 0.85
Feedback-prop $(B\|A)$	3.81	**1.54**	52.51 ± 0.2	83.97 ± 0.37	35.83 ± 0.95
R-Tran $(B\|A)$	**3.6**	1.56	**52.82** ± 0.78	**84.21** ± 0.63	**36.31** ± 0.8

- **R-Tran (ours)** outperforms other methods in predicting species encounter rates with partial information both within a taxon (birds) and across taxa (birds and butterflies).