



LifeCLEF 2022 Lab: multimedia species identification challenges

Alexis Joly, Titouan Lorieul, Stefan Kahl, Lukáš Pícek, Hervé Goëau, Elijah Cole, Benjamin Deneu, Maximilien Servajean, Andrew Durso, Isabelle Bolon, Hervé Glotin, Robert Planqué, Willem-Pier Vellinga, Amanda Navine, Holger Klinck, Tom Denton, Ivan Eggel, Pierre Bonnet, Milan Šulc, Henning Müller



PlantCLEF: image-based plant identification at global level

Scenario: Scaling-up the number of species identified by Photo ID applications such as Pl@ntNet. A step towards covering the entire earth's flora (80K species among 300K on earth).



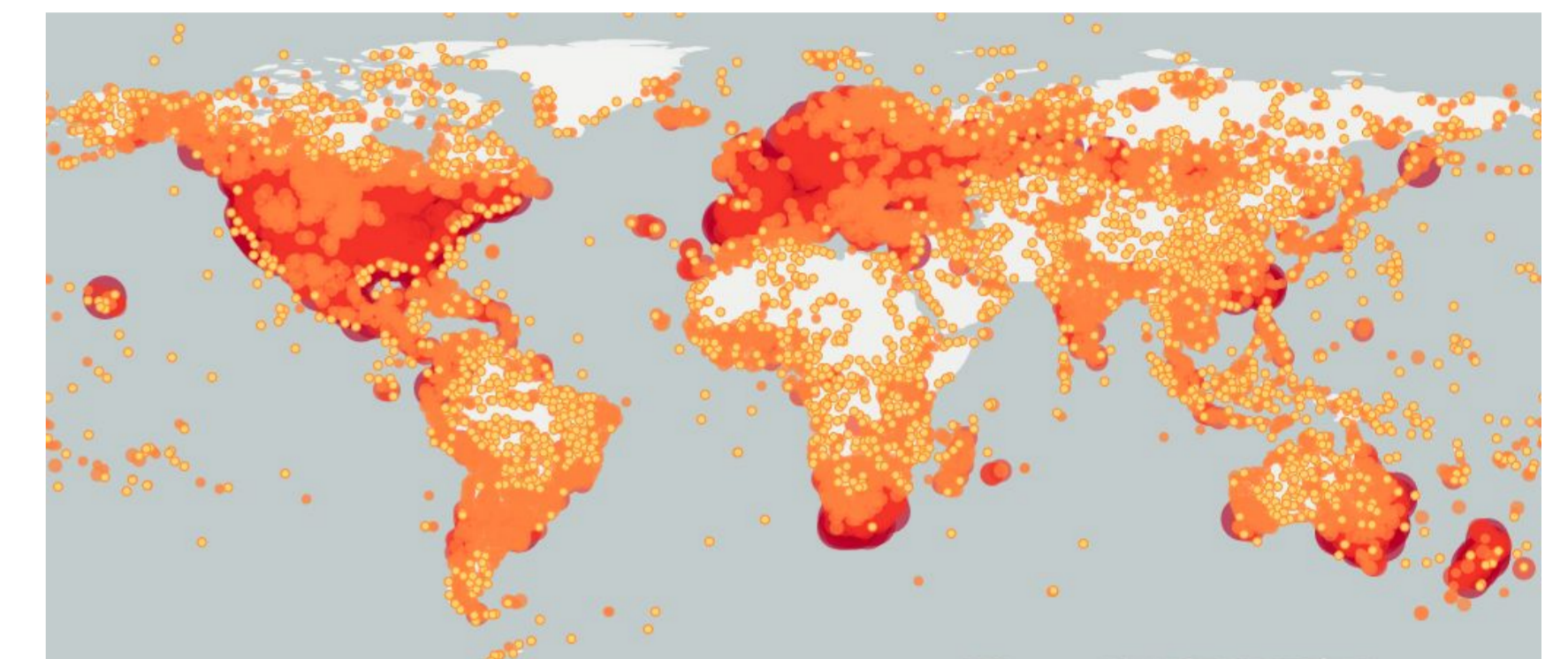
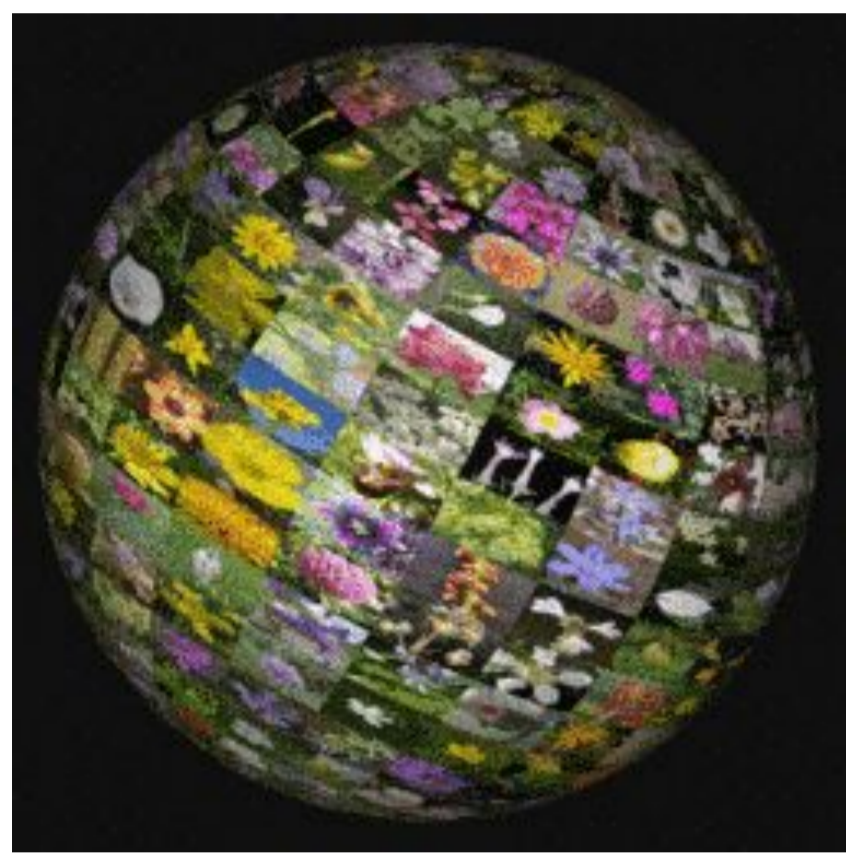
Task: multi-image classification (80K classes, 1 to 5 images per test sample)

Data:

- Training set = 4M images categorized in 80k classes
 - Two sources: 2.9M from GBIF (biodiversity platform), 1.1M from the web (noisy)
 - Severely imbalanced (long-tail distribution)
- Test set = 26K plant observations coming from Pl@ntNet (1 to 5 images per observation)

Metric: Macro-averaged Mean Reciprocal Rank (inverse of the rank of the correct answer)

Competition: PlantCLEF 2022 is held on AICrowd as a competition with hidden test data. Participants are required to submit working notes to be published.



BirdCLEF: identifying bird calls in soundscape recordings

Scenario: As the "extinction capital of the world," Hawai'i has lost 68% of its bird species, the consequences of which can harm entire food chains. BirdCLEF 2022 focuses on the automated detection of native Hawai'ian bird species in soundscape recordings.

Task: Identify all vocalizing birds within a 5-second segment of audio data.

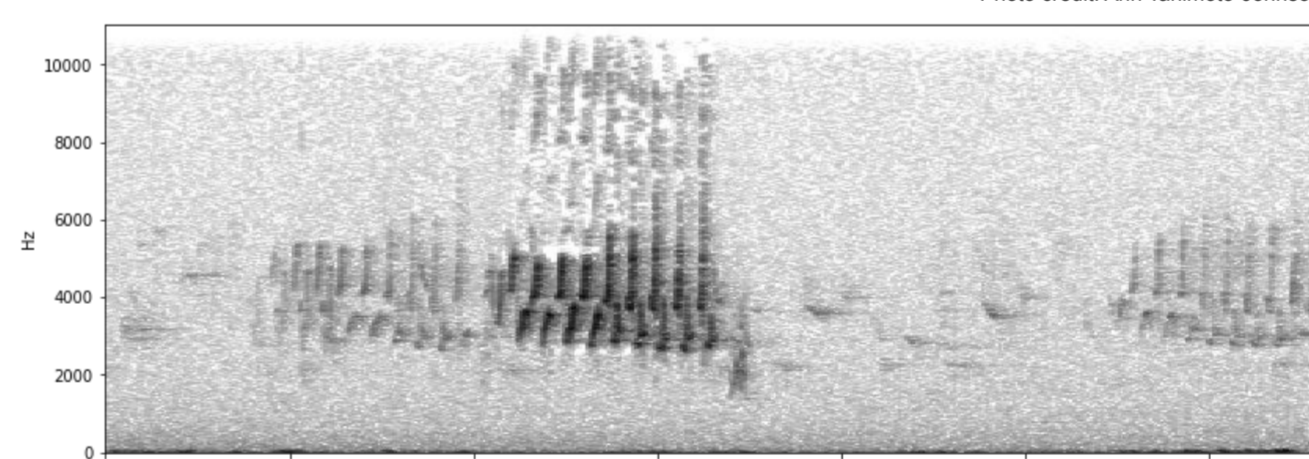
- Data:**
- ~15,000 audio recordings from Xeno-canto covering 152 species for training
 - ~90 hrs of annotated soundscape recordings as test data

Metric: Macro-averaged F1-score

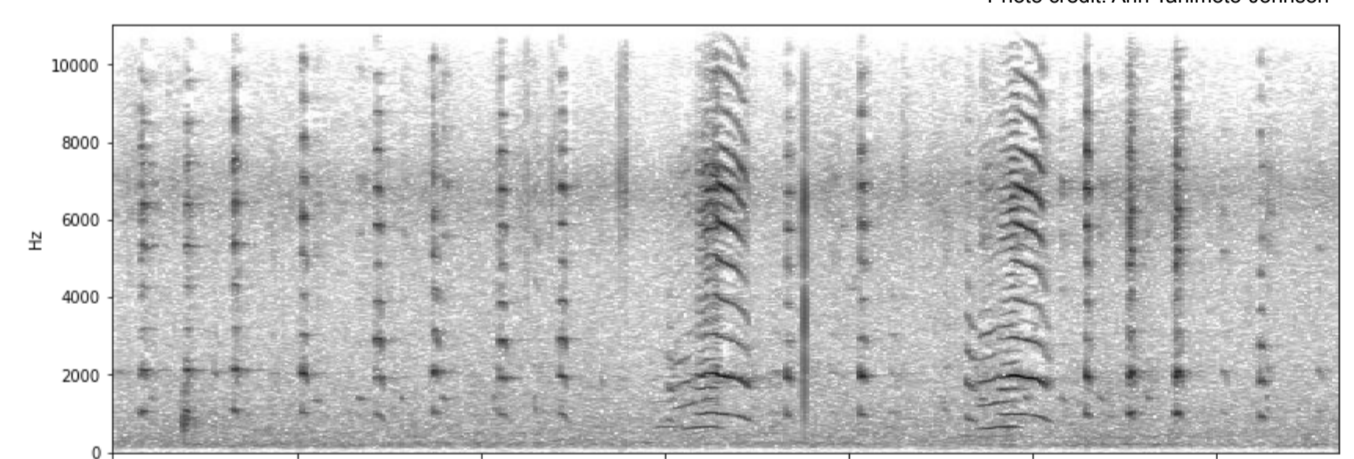
Competition: BirdCLEF 2022 is held on Kaggle as a code competition with hidden test data. Participants are encouraged to submit working notes.

Prizes: \$5,000 split between the three best performing systems and \$5,000 in Google Cloud credits split between the two best working notes.

Kiwikiu (*Pseudonestor xanthophrys*)



Nēnē (*Branta sandvicensis*)



More information online at: <https://www.kaggle.com/c/birdclef-2022>

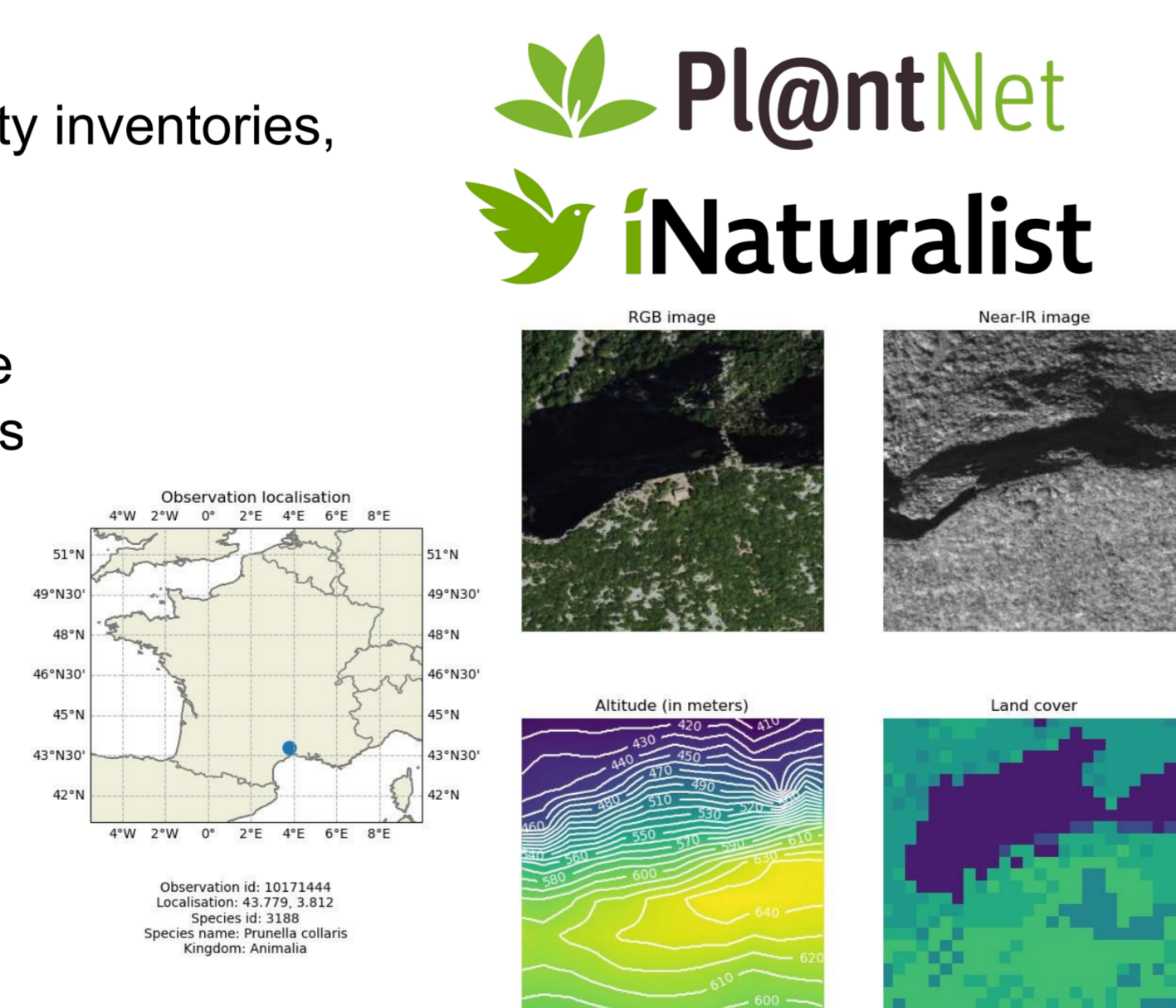
GeoLifeCLEF: location-based species presence prediction

Task: Predict the list of plant and animal species that are the most likely to be observed at a given location in France and US

Scenario: invasive species detection, facilitate biodiversity inventories, prediction of new potential habitats, etc.

- Data:**
- 1.6M observations of 17K species from citizen science
 - 1 observation = 1 species name + GPS coordinates
 - For each location:
 - High-resolution patches (1px = 1m)
 - RGB + Near-IR aerial images, altitude and land cover data
 - Low-resolution rasters
 - climatic and pedologic data

Evaluation: top-30 error rate of predicted lists of species averaged over test observations



SnakeCLEF: identification in medically important scenarios

Task: Given the set of snake observations – multiple photographs of the same individual – and corresponding geographic locality information, the goal of the task is to return a species ID to all the observations.

Data: 187,129 snake observations with 318,532 photographs belonging to 1,572 snake species and observed in 208 countries. The data were gathered from the online biodiversity platform – iNaturalist.

The provided dataset has a heavy long-tailed class distribution, where the most frequent species (Natrix natrix) is represented by 6,472 images and the least frequent species by just 5 samples.

Main novelties over 2021:

- Reduced bias towards North America and Europe.
- Human vs Machine Experiment.
- Evaluation beyond the main challenge metrics (Mean-F1 Score):
 - "Harmless vs. Venomous species".

Challenge: <https://www.kaggle.com/c/snakeclef2022>



FungiCLEF: Fungi recognition as Open-set problem

Task: The goal of the task is to predict a fungi species (known or "unknown") for each observation based on a given set of images and appropriate metadata.

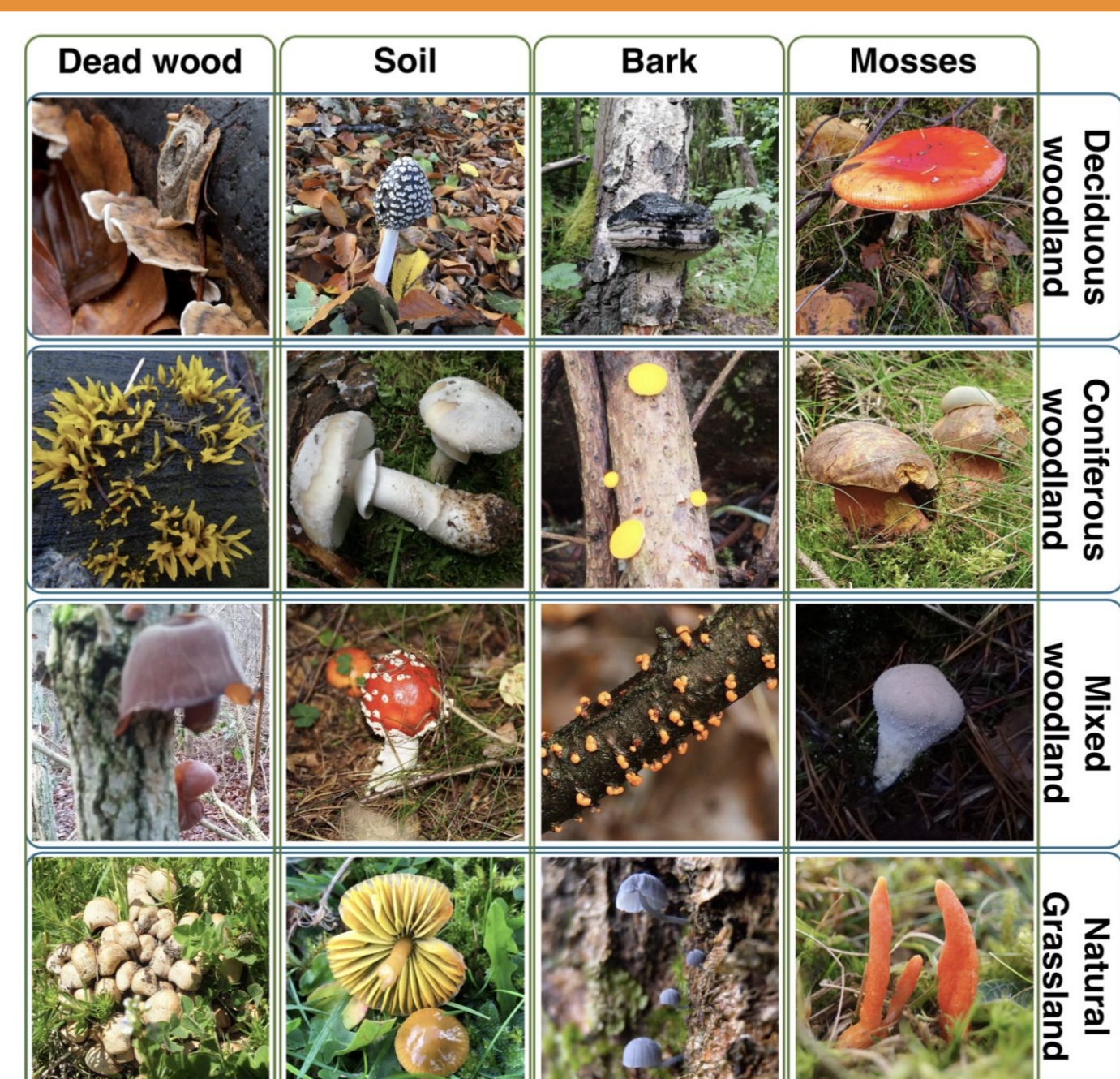
Data: 295,938 training images belonging to 1,604 species observed mostly in Denmark. All training samples passed an expert validation process, guaranteeing high-quality labels.

The test set contains 59,420 observations with 118,676 images and 3,134 species. We include observations collected through the whole 2021 and across all substrate and habitat types.

Additional perks:

- Images and rich metadata for all observations: Habitat, Substrate, Time, Location, Device, etc.
- Evaluation beyond the main challenge metrics (Mean-F1 Score): "Edible vs. poisonous".
- Open-set classification task

Challenge link: <https://www.kaggle.com/c/fungiclef2022>



Registration / participation

CLEF registration: LifeCLEF lab is part of the Conference and Labs of the Evaluation Forum (CLEF) that includes a larger set of benchmarking activities carried within various virtual labs. Registration to CLEF labs is free: <https://clef2022-labs-registration.dei.unipd.it/>

Participation: Participation to BirdCLEF, GeoLifeCLEF and SnakeCLEF is done through Kaggle challenges whereas participation to PlantCLEF and SnakeCLEF is done through AICrowd challenges. The exact URL of each challenge can be found on LifeCLEF website (www.lifeclef.org).

Working notes: Methods and results of participants are published in working notes, i.e. technical papers published within CEUR-WS proceedings (with an individual DOI and an indexing by many bibliography systems including DBLP, Google scholar, etc.).

CLEF 2022 conference and LifeCLEF workshop: A one-day workshop dedicated to LifeCLEF will be held within CLEF 2022 conference (5-8 September 2022, Bologna, Italy). The programme will include (i) the presentation of the main outcomes by the organizers, (ii) some participant's talks, and (iii) invited speakers at the crossroads of machine learning, ecology and citizen sciences.