LifeCLEF 2019: Biodiversity Identification and Prediction Challenges

Alexis Joly, Hervé Goëau, Christophe Botella, Stefan Kahl, Marion Poupard, Maximilien Servajean, Hervé Glotin, Pierre Bonnet, Willem-Pier Vellinga, Robert Planqué, et al.

LifeCLEF 2019: Biodiversity Identification and Prediction Challenges

Alexis Joly[1], Hervé Goëau[2], Christophe Botella[1,3], Stefan Kahl[7], Marion Poupard[4], Maximillien Servajean[8], Hervé Glotin[4], Pierre Bonnet[2], Willem-Pier Vellinga[5], Robert Planqué[5], Jan Schlüter[4], Fabian-Robert Stöter[1], Henning Müller[6]

[1] Inria, LIRMM, Montpellier, France
[2] CIRAD, UMR AMAP, France
[3] INRA, UMR AMAP, France
[4] AMU, Univ. Toulon, CNRS, ENSAM, LSIS UMR 7296, IUF, France
[5] Xeno-canto foundation, The Netherlands
[6] HES-SO, Sierre, Switzerland
[7] Chemnitz University of Technology, Germany
[8] LIRMM, Université Paul Valéry, University of Montpellier, CNRS, Montpellier, France

Abstract. Building accurate knowledge of the identity, the geographic distribution and the evolution of living species is essential for a sustainable development of humanity, as well as for biodiversity conservation. However, the burden of the routine identification of plants and animals in the field is strongly penalizing the aggregation of new data and knowledge. Identifying and naming living plants or animals is actually almost impossible for the general public and often a difficult task for professionals and naturalists. Bridging this gap is a key challenge towards enabling effective biodiversity information retrieval systems. The LifeCLEF evaluation campaign, presented in this paper, aims at boosting and evaluating the advances in this domain since 2011. In particular, the 2019 edition proposes three data-oriented challenges related to the identification and prediction of biodiversity: (i) an image-based plant identification challenge, (ii) a bird sounds identification challenge and (iii) a location-based species prediction challenge based on spatial occurrence data and environmental tensors.

Keywords: biodiversity, informatics, machine learning, species identification, species prediction, plant identification, bird identification, species distribution model

1 Introduction

Identifying organisms is a key for accessing information related to the uses and ecology of species. This is an essential step in recording any specimen on earth to be used in ecological studies. Unfortunately, this is difficult to achieve due

to the level of expertise necessary to correctly record and identify living organisms (for instance plants are one of the most difficult groups to identify with an estimated number of 400,000 species). This *taxonomic gap* has been recognized since the Rio Conference of 1992, as one of the major obstacles to the global implementation of the Convention on Biological Diversity. Among the diversity of methods used for species identification, Gaston and O'Neill[2] discussed in 2004 the potential of automated approaches typically based on machine learning and multimedia data analysis. They suggested that, if the scientific community is able to (i) overcome the production of large training datasets, (ii) more precisely identify and evaluate the error rates, (iii) scale up automated approaches, and (iv) detect novel species, it will then be possible to initiate the development of a generic automated species identification system that could open up vistas of new opportunities for theoretical and applied work in biological and related fields.

Since the question raised by Gaston and O'Neill[2], *automated species identification: why not?*, a lot of work has been done on the topic (*e.g.* [13, 1, 17, 16, 4, 5, 11]) and it is still attracting much research today, in particular in deep learning [3, 6, 14]. In order to measure the progress made in a sustainable and repeatable way, the LifeCLEF[9] research platform was created in 2014 as a continuation of the plant identification task [10] that was run within the ImageCLEF lab [10] the three years before [8, 9, 7]. LifeCLEF enlarged the evaluated challenge by considering animals in addition to plants, and audio and video contents in addition to images. In 2018, a new challenge dedicated to the location-based prediction of species was finally introduced (GeoLifeCLEF).

2   PlantCLEF Challenge

2.1   Methodology

The plant identification challenge of CLEF has been run since 2011, offering today a seven-year follow-up of the progress made in image-based plant identification. From the beginning, it mainly relied on real-world collaborative data and the evaluation protocol was defined in collaboration with biologists so as to reflect realistic usage scenarios. In particular, it considers the problem of classifying plant observations based on several images of the same individual plant rather than considering a classical image classification task. Indeed, it is usually required to observe several organs of a plant to identify it accurately (*e.g.* the flower, the leaf, the fruit, the stem, etc.). As a consequence, the same individual plant is often photographed several times by the same observer resulting in contextually similar pictures and/or near-duplicates. To avoid bias, it is crucial to consider such image sets as a single plant observation that should not be split across the training and the test set. In addition to the raw pictures, plant observations are usually associated with contextual and social data. This

---

[9] http://www.lifeclef.org/
[10] http://www.imageclef.org/

includes geo-tags or location names, time information, author names, collaborative ratings, vernacular names (common names), picture type tags, etc. Within all PlantCLEF challenges, the use of this additional information was considered as part of the problem because it was judged as potentially useful for a real-world usage scenario.

The data that was shared within the PlantCLEF challenge was considerably enriched along the years. The number of species was increased from 71 species in 2011 to 10,000 species in 2017 and 2018 (illustrated by more than 1 million images). This durable scaling-up was made possible thanks to the close collaboration of LifeCLEF with several important actors in the digital botany domain, in particular the TelaBotanica network of expert and amateur botanists (about 40K members) and the Pl@ntNet citizen science platform (million of users).

2.2   Main Outcomes of the Previous Edition

The main novelty of the 2018 edition of PlantCLEF was to involve 9 of the best expert botanists of the French flora who accepted to compete with AI algorithms on a difficult subset of the whole test set. The results confirmed that identifying plants from images is a difficult task, even for some of the highly skilled specialists who accepted to participate in the experiment. Images only contain a partial information of the plant and that it is often not sufficient to determine the right species with certainty. Regarding the performance of the automated approaches, the results showed that there is still a margin of progression but that it is becoming tighter and tighter. The best system was able to correctly classify 84% of the test samples, better than 5 of the 9 experts.

2.3   PlantCLEF 2019

The main novelty of the 2019 edition of PlantCLEF will be to extend the challenge to the flora of *data deficient* regions, *i.e.* regions having the richest biodiversity (tropical ones) but for which data availability is much lower than northern countries. Indeed, it is estimated that there is over 391K species of vascular plants on earth, much beyond the 10K species of PlantCLEF 2018 that are among the most common ones. The additional data will be aggregated in two ways. For the training set, we will mainly rely on raw web data collected by querying popular image search engines with the binomial latin name of the targeted species. We actually did show in previous editions of LifeCLEF that training deep learning models on such noisy big data is as effective as training models on cleaner but smaller expert data. For the test set, on the other hand, we will use expert data without any uncertainty. More precisely, we will rely on 3 collections of expert botanists who accepted to share their unpublished observations for the challenge. One is a collection of trees, shrubs, herbs and ferns from French Guyana (wet evergreen Amazonian forest). The second one is a specialized collection of pictures related to epiphytic orchids, mainly from Laos. And the third one is a collection of endemic species of South Africa. The main evaluation measure for the challenge will be the Mean Reciprocal Rank.

3   BirdCLEF Challenge

3.1   Methodology

The bird identification challenge of LifeCLEF, initiated in 2014 in collaboration with Xeno-Canto, considerably increased the scale of the seminal challenges. The first bird challenge ICML4B [4] initiated in 2012 by DYNI/SABIOD had only 35 species, but received 400 runs. The next at MLSP had only 15 species, the third (NIPS4B [5] in 2013 by SABIOD) had 80 species. Meanwhile, Xeno-canto, launched in 2005, hosts bird sounds from all continents and daily receives new recordings from some of the remotest places on Earth. It currently archives with 379472 recordings, 9779 species of birds, making it one of the most comprehensive collections of bird sound recordings worldwide, and certainly the most comprehensive collection shared under Creative Commons licenses. For the first Bird-CLEF challenge, it was decided to not consider the whole Xeno-Canto dataset but to rather focus on a specific region, *i.e.* the Amazonian rain forest because it is one of the richest in the world in terms of biodiversity but also one of the most endangered. The geographical extent and the number of species were progressively increased over the years so as to reach 1000 species in 2015/2016, and 1500 in 2017/2018. By nature, the Xeno-Canto data as well as the BirdCLEF subset has a massive class imbalance. For instance, the 2017 dataset contains 48,843 recordings in total, with a minimum of four recordings for *Laniocera rufescens* and a maximum of 160 recordings for *Henicorhina leucophrys*.

In 2016, the BirdCLEF challenge was extended to *soundscape* recordings in addition to the classical *mono-directional* Xeno-Canto recordings. This enables more passive monitoring scenarios such as setting up a network of static recorders that would continuously capture the surrounding sound environment. One of the limitations of this new content, however, was that the vocalizing birds were not localized in the recordings. Thus, to allow a more accurate evaluation, new *time-coded soundscapes* were introduced within the BirdCLEF 2017 and 2018 challenges. In total, 6.5 hours of recordings were collected in the Amazonian forests and were manually annotated by two experts including a native of the Amazon forest, in the form of time-coded segments with associated species name.

3.2   Main Outcomes of the Previous Edition

The best system of the 2018 edition of the BirdCLEF challenge achieved an impressive Mean Average Precision score of 0.83 on the mono-directional recordings. This performance could probably even be improved by a few points by combining it with a metadata-based prediction model, as shown by the second best participant to the challenge. This means that the technology is now mature enough for this scenario. Concerning the soundscapes recordings however, we did not observe any significant improvement over the performance of the 2017 edition. Recognizing many overlapping birds remains a hard problem and none of the efforts made by the participants to tackle it provided observable improvement.

3.3   BirdCLEF 2019

The 2019 edition of the BirdCLEF challenge will mainly focus on the soundscape scenario that remains very challenging whereas the mono-directional identification task is now better solved. Two tasks will be evaluated, (i) the recognition of all specimens singing in a long sequence (up to one hour) of raw soundscapes that can contain tens of birds singing simultaneously, and (ii) source separation or source count estimation in complex soundscapes that were recorded using multiple microphones. Therefore, two new corpus of soundscapes will be added to the existing soundscape dataset: (i) 100+ hours of manually annotated soundscapes recorded using 30 field recorders between January and June of 2017 in Ithaca, NY, USA. (ii) 50 hours of four-channel or stereophonic binaural recordings acquired in Papa New Guinea in november 2017 at high sampling rate (96 kHz SR) and high dynamics (24 bits) [15]. For this purpose we designed binaural or quadriphonic recording stations, specifically for localisation in azimuth and elevation of singing birds, in order to help in a second stage the recognition of the species. These recordings contain some endemic bird species that had never been recorded before. The evaluation measure used for the species detection task will be the classification mean Average Precision (c-mAP [12]). The evaluation measure used for the count estimation task is the mean absolute count error.

4   GeoLifeCLEF Challenge

4.1   Methodology

Predicting the shortlist of species that are likely to be observed at a given geographical location should significantly help to reduce the candidate set of species to be identified. However, none of the attempt to do so within previous Life-CLEF editions successfully used this information. The GeoLifeCLEF challenge was specifically created in 2018 to tackle this problem through a standalone task. More generally, automatically predicting the list of species that are likely to be observed at a given location might be useful for many other scenarios in biodiversity informatics. It could facilitate biodiversity inventories through the development of location-based recommendation services (typically on mobile phones) as well as the involvement of non-expert nature observers. It might also serve educational purposes thanks to biodiversity discovery applications providing functionalities such as contextualized educational pathways.

The challenge relies on a large data set of 291,392 occurrences of around 3K plant species, each occurrence being associated to a location, a species name and a multi-channel image characterizing the local environment. Indeed, it is usually not possible to learn a species distribution model directly from spatial positions because of the limited number of occurrences and the sampling bias. What is usually done in ecology is to predict the distribution on the basis of a representation in the environmental space, typically a feature vector composed of climatic variables (average temperature at that location, precipitation, etc.) and other variables such as soil type, land cover, distance to water, etc. The

originality of GeoLifeCLEF is to generalize such niche modeling approach to the use of an image-based environmental representation space. Instead of learning a model from environmental feature vectors, the goal of the task will be to learn a model from k-dimensional image patches, each patch representing the value of an environmental variable in the neighborhood of the occurrence.

4.2   Main Outcomes of the Previous Edition

The main outcome of the first edition of GeoLifeCLEF was that Convolutional Neural Networks models learned on environmental tensors revealed to be the most performing method. They performed better than boosted classification trees that were known as providing state-of-the-art performance for environmental modelling. However, the achieved performance is still low with regard to the targeted scenario and there is a large room of improvement and research opportunities regarding such models, like appropriately integrating neighbours species correlations in the model, using external expert information about related species like taxonomic or phylogenetic classification, or correcting for observer reporting bias.

4.3   GeoLifeCLEF 2019

The 2019 edition of the challenge will tackle some of the methodological weaknesses that were revealed by the pilot 2018 edition. In particular, we will rely on the top-30 accuracy instead of the Mean Average Precision as the main evaluation metric. This will allow to better take into account the fact that many species co-exist at small spatial scales (under the meter), much lower than the accuracy of the geo-coordinates in the data set. We will also produce a new dataset fixing some issues of the previous one related to the incompleteness of some environmental variables and the spatial degradation of some occurrences. More precisely, the training set will be composed of nearly one million geo-locations of plant species living on the French territory (coming from two main platforms: (i) the Global Biodiversity Information Facility and (ii), the Pl@ntNet participatory application). For the test set, on the other hand, we will rely solely on expert data without any uncertainty coming from the French national conservatories. Regarding the environmental variables, we will provide about 30 rasters of data covering the whole French territory (related to climatology, altitude, soil type, land cover, distance to water, etc.). We will also provide tools to extract environmental tensors from that rasters (at the positions of the plant occurrences in the training and test sets).

5   Timeline and registration instructions

All information about the timeline and the participation to the challenges is provided on the LifeCLEF 2019 web pages[11]. The system used to run the challenges (registration, submission, leaderboard, etc.) is the crowdAI platform[12].

6   Discussion and conclusion

Boosting research on biodiversity informatics in the long term is crucial in terms of societal impact. Researchers are actually often opportunistic regarding the choice of a dataset and an interesting related challenge. And so are end-users regarding the use of applications emerging from that research. To fully reach its objective, an evaluation campaign such as LifeCLEF requires a long-term research effort so as to (i) encourage non-incremental contributions, (ii) measure consistent performance gaps, (iii) progressively scale-up the problem and (iv), enable the emergence of a strong community. The 2019-th edition of the lab will support this vision but will still include a set of consistent novelties:

 – The historical BirdCLEF subtask related to monospecies recordings will be stopped in order to concentrate all efforts on the most challenging subtask of recognizing birds in soundscapes and on a new subtask relying on polyphonic recordings.
 – We will go deeper in the comparison of automated approaches with human expertise by extending the PlantCLEF task to more complex taxonomic groups, in particular the floras of several tropical countries that are known only by a few specialists who will participate in the evaluation.
 – The evaluation methodology of the GeoLifeCLEF challenge will be improved according to the feedback of the first edition and the dataset will be enriched with more diverse and more precise plant occurrences.

References

1. Cai, J., Ee, D., Pham, B., Roe, P., Zhang, J.: Sensor network for the monitoring of ecosystem: Bird species recognition. In: Intelligent Sensors, Sensor Networks and Information, 2007. ISSNIP 2007. 3rd International Conference on (2007). https://doi.org/10.1109/ISSNIP.2007.4496859
2. Gaston, K.J., O'Neill, M.A.: Automated species identification: why not? Philosophical Transactions of the Royal Society of London B: Biological Sciences **359**(1444), 655–667 (2004)
3. Ghazi, M.M., Yanikoglu, B., Aptoula, E.: Plant identification using deep neural networks via optimization of transfer learning parameters. Neurocomputing **235**, 228–235 (2017)

---

[11] https://www.imageclef.org/lifeclef2019
[12] https://www.crowdai.org

4. Glotin, H., Clark, C., LeCun, Y., Dugan, P., Halkias, X., Sueur, J.: Proc. 1st workshop on Machine Learning for Bioacoustics - ICML4B. ICML, Atlanta USA (2013)
5. Glotin, H., LeCun, Y., Artiéres, T., Mallat, S., Tchernichovski, O., H.X.: Proc. Neural Information Processing Scaled for Bioacoustics, from Neurons to Big Data. NIPS Int. Conf., Tahoe USA (2013), http://sabiod.org/nips4b
6. Goeau, H., Bonnet, P., Joly, A.: Plant identification based on noisy web data: the amazing performance of deep learning (lifeclef 2017). In: CLEF 2017-Conference and Labs of the Evaluation Forum. pp. 1–13 (2017)
7. Goëau, H., Bonnet, P., Joly, A., Bakic, V., Barthélémy, D., Boujemaa, N., Molino, J.F.: The imageclef 2013 plant identification task. In: CLEF. Valencia, Spain (2013)
8. Goëau, H., Bonnet, P., Joly, A., Boujemaa, N., Barthélémy, D., Molino, J.F., Birnbaum, P., Mouysset, E., Picard, M.: The imageclef 2011 plant images classification task. In: CLEF 2011 (2011)
9. Goëau, H., Bonnet, P., Joly, A., Yahiaoui, I., Barthélémy, D., Boujemaa, N., Molino, J.F.: Imageclef2012 plant images identification task. In: CLEF 2012. Rome (2012)
10. Goëau, H., Joly, A., Bonnet, P., Bakic, V., Barthélémy, D., Boujemaa, N., Molino, J.F.: The imageclef plant identification task 2013. In: Proceedings of the 2nd ACM international workshop on Multimedia analysis for ecological data. pp. 23–28. ACM (2013)
11. Joly, A., Goëau, H., Bonnet, P., Bakić, V., Barbe, J., Selmi, S., Yahiaoui, I., Carré, J., Mouysset, E., Molino, J.F., et al.: Interactive plant identification based on social image data. Ecological Informatics **23**, 22–34 (2014)
12. Joly, A., Goëau, H., Botella, C., Glotin, H., Bonnet, P., Vellinga, W.P., Planqué, R., Müller, H.: Overview of lifeclef 2018: a large-scale evaluation of species identification and recommendation algorithms in the era of ai. In: International Conference of the Cross-Language Evaluation Forum for European Languages. pp. 247–266. Springer (2018)
13. Lee, D.J., Schoenberger, R.B., Shiozawa, D., Xu, X., Zhan, P.: Contour matching for a fish recognition and migration-monitoring system. In: Optics East. pp. 37–48. International Society for Optics and Photonics (2004)
14. Lee, S.H., Chan, C.S., Remagnino, P.: Multi-organ plant classification based on convolutional and recurrent neural networks. IEEE Transactions on Image Processing **27**(9), 4287–4301 (2018)
15. Poupard, M., Glotin, H., Lengagne, T., Bougrain-Dubourg, A., Bedu, A.L.: Multichannel soundscape recordings of wild versus anthropised new guinea - a first acoustic repertory of some endemic species. LIS DYNI CNRS Toulon Research Report (2018)
16. Towsey, M., Planitz, B., Nantes, A., Wimmer, J., Roe, P.: A toolbox for animal call recognition. Bioacoustics **21**(2), 107–125 (2012)
17. Trifa, V.M., Kirschel, A.N., Taylor, C.E., Vallejo, E.E.: Automated species recognition of antbirds in a mexican rainforest using hidden markov models. The Journal of the Acoustical Society of America **123**, 2424 (2008)