Implicit biodiversity monitoring from mobile search logs

Alexis Joly
Inria, ZENITH team, LIRMM
34095 Montpellier, France

Hervé Goëau
IRD, UMR AMAP
Montpellier, France

Julien Champ
Inria, ZENITH team, LIRMM
34095 Montpellier, France

Samuel Dufour-Kowalski
INRA, UMR AMAP
Montpellier, France

Henning Müller
HES-SO
3960 Sierre, Switzerland

Pierre Bonnet
CIRAD, UMR AMAP
Montpellier, France

ABSTRACT

Large scale biodiversity monitoring is essential for sustainable development (earth stewardship). With the recent advances in computer vision, we see the emergence of more and more effective identification tools allowing to set-up large-scale data collection platforms such as the popular Pl@ntNet initiative. Although it still covers only a fraction of the world flora, this platform is already being used by more than 300K people who produce tens of thousands of validated plant observations each year. Nevertheless, this explicitly shared and validated data is only the tip of the iceberg. The real potential relies on the millions of raw image queries submitted by the users of the mobile application but for which there is no human validation at all. Allowing the exploitation of such contents in a fully automatic way could scale up the world-wide collection of plant observations by several orders of magnitude. In this paper, we first survey existing automated plant identification systems through a five-year synthesis of the PlantCLEF benchmark and an impact study of the Pl@ntNet platform. We then focus more specifically on the implicit monitoring scenario and discuss several new related research challenges. Finally, we discuss the results of a preliminary experimental study focused on the implicit monitoring of invasive species in mobile search logs. We show that the results are very promising but that there is still some room for improvement before being able to automatically share such implicit observations within international biodiversity platforms.

1. INTRODUCTION

Identifying organisms is a key step in accessing information related to the ecology of species. This is an essential step in recording any specimen on earth to be used in ecological studies. But unfortunately, this is difficult to achieve due to the level of expertise necessary to correctly identify and record living organisms (in particular plants that are one of the most difficult group to identify with more than 300,000 species on earth). 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 et al.[11] discussed in 2004 the potential of automated approaches typically based on machine learning and multimedia data analysis methods. 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 pure and applied work in biological and related fields.

Since the question raised by Gaston in 2004 ("automated species identification: why not?"), enormous work has been done on the development of automated approaches for plant species identification, mostly based on computer vision techniques (*e.g.* [4, 17, 46, 23, 24, 28, 45]). Some of these results have been integrated in effective web or mobile tools and have initiated close interactions between computer scientists and end-users such as ecologists, botanists, educators, land managers and the general public. One of the first remarkable system in this domain was the LeafSnap application [27], focused on a few hundreds tree species of North America. This was followed few years later by other applications such as Pl@ntNet [22] or Folia [6] more specifically dedicated to the European flora, or LikeThat garden[1] more focused on garden plants. These productions were perceived as innovative tools and have received a good support of a large part of the society. The number of news articles on the web dedicated to this subject is a good illustration of this positive perception. These tools are nevertheless at their early stage of development according to the large number of plant species on earth, the large diversity of end-users interested in such an accessible approach and the limits of today's performance.

In parallel to the emergence of automated identification tools, large social networks dedicated to the production, sharing and identification of biodiversity records have increased in recent years. Some of the most active ones in the botanical domain like iNaturalist[2], iSpot [35], Tela Botan-

[1] https://www.likethatapps.com/LikeThatGarden/
[2] http://www.inaturalist.org/

ica[3], respectively initiated in the US for the two first ones and in Europe for the last one, federate tens of thousands of members, producing hundreds of thousands of observations. As a proof of their increasing reliability, some of these networks started to contribute to global initiatives in biodiversity such as the Global Biodiversity Information Facility (GBIF[4]) which is the largest and most recognized one.
Noticeably, the Pl@ntNet initiative was the first one attempting to combine the force of social networks with that of automated identification tools [23]. It was launched in 2009 by a trans-disciplinary French consortium involving research institutes in computer sciences, ecology and agriculture in collaboration with the Tela Botanica social network. This was the starting point of several scientific and technological productions [12] which finally led to the first release of the Pl@ntNet app (iOS in February 2013 [16] and Android [15] the following year). It was the first system allowing the use of a combination of different visual features (such as leaf, stem, fruit and flower) and the first one relying on a continuously enriched collaborative training set. This app was initially based on 800 species and was progressively enlarged to thousands of plant species of the European region (6 140 species up to now). Nowadays, the platform is being used by about 300K people who produce tens of thousands of validated plant observations each year thanks to collaborative validation tools (IdentiPlante[5] and PictoFlora[6]). Most of these observations (actually the geo-localized ones) contribute to the global knowledge of plant species thanks to their publication by the GBIF.
Nevertheless, this explicitly shared and validated data is only the tip of the iceberg. The real potential relies in the millions of raw image queries submitted by the users of the mobile application but for which there is no human validation at all. As an illustration, in 2015, 2,328,502 queries have been submitted by the users of the Pl@ntNet mobile apps but only less than 1 % of them have been finally shared and collaboratively validated. Allowing the exploitation of the unvalidated observations in a fully automatic way could scale up the world-wide collection of plant records by several orders of magnitude. More generally, the idea of implicitly monitoring living organisms from any kind of User Generated Content data streams has the potential to revolutionize biodiversity monitoring at a very limited cost.

In this paper, we first survey existing systems and approaches for the automated collection of plant observations. Next, we provide a five-year overview of the PlantCLEF international benchmark (organized since 2011 within the ImageCLEF and LifeCLEF events) as well as the results of an impact study of the Pl@ntNet infrastructure that we conducted through a survey of more than 700 respondents. We then specifically address the scientific and technological challenges related to the implicit monitoring scenario. We show that it is indeed highly related to several hard problems such as to novelty and uncertainty, and we suggest brave new research perspectives to address them. Finally, as a first concrete step, we provide the results of an experimental study focused on the implicit monitoring of invasive species (*i.e.* an alien species of which the introduction does or is likely to cause economic or environmental harm or harm to human health) in Pl@ntNet mobile search logs (that was part of the LifeCLEF 2016 evaluation campaign).

[3] `http://www.tela-botanica.org/`
[4] `http://www.gbif.org/`
[5] `http://www.tela-botanica.org/appli:identiplante`
[6] `http://www.tela-botanica.org/appli:pictoflora`

2. THE PLANTCLEF CHALLENGE: A FIVE-YEAR OVERVIEW

In order to evaluate the performance of automated plant identification technologies in a sustainable and repeatable way, a dedicated system-oriented benchmark was setup in 2011 in the context of the international evaluation campaign ImageCLEF[7]. In 2011, 2012 and 2013 respectively 8, 11 and 12 international research groups participated in this large collaborative evaluation by benchmarking their image-based plant identification systems (see [18, 19, 17] for more details). It 2014, the LifeCLEF[8] research platform was created in the continuity of this effort so as to enlarge the evaluated challenges by considering birds and fishes in addition to plants, and audio and video contents in addition to images.
Within this context, the plant identification benchmark continued to be run yearly offering today a five-year follow-up of the progress in image-based plant identification. A particularity of the benchmark is that it always focused on real-world collaborative data contrary to most other testbeds found in the literature that were created through well controlled laboratory conditions. Additionally, the evaluation protocol was defined in collaboration with biologists and environmental stakeholders so as to reflect realistic usage scenarios. Notably particular attention was accorded to the notion of observation rather than considering standalone images. In practice, the same individual plant is actually 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 set as a single plant observation that should not be split across the training and test set. Besides this, the use of contextual and social data was also authorized when they were judged as potentially useful and accessible in a real-world usage scenario. This includes geo-tags or location names, time information, author names, collaborative ratings, vernacular names (common names), picture type tags, etc. It is however important to note that the visual modality remained largely predominant in all the best systems along the years and that the use of metadata was shown to provide only slight additional improvements.

Tables 1 and 2 give a year-to-year overview of the shared data and of the best performing systems (detailed descriptions of the results and systems can be found in the technical overview papers of each year [18, 19, 17, 20, 13, 14] and participant working notes). To allow a comprehensive comparison along the years, we isolated in Table 1 the *leaf scans* and *white background* image categories that were part of the evaluation of the three first years but that were abandoned afterwards. Table 2, on the other side, focuses on photographs of plants in their natural environment (only leaves in 2011-2012, diverse organs and plant views in the following years). For a fair comparison, we also removed from the overview, the submissions that were humanly assisted in some point (*e.g.* involving a manual segmentation

[7] `www.imageclef.org`
[8] `www.lifeclef.org`

Table 1: Three-year synthesis of the PlantCLEF challenge restricted to *leaf scans* and *pseudo-scans*

Year	#Species	#Images	#Evaluated systems	Score of best system	Brief description of best system
2011	71	3,967	20	0.574	. Various local features (around Harris points) . Hash-based indexing . RANSAC based matching
2012	126	9,356	30	0.565	. Shape and texture global features . SVM classifier
2013	250	11,031	33	0.607	. Shape and texture global features . SVM classifier

of the leaves). The evaluation metric that was used from 2011 to 2015 was *i.e.* an extension of the mean reciprocal rank [41] classically used in information retrieval. The difference is that it is based on a two-stage averaging rather than a flat averaging such as:

$$S = \frac{1}{U}\sum_{u=1}^{U}\frac{1}{P_u}\sum_{p=1}^{P_u}\frac{1}{r_{u,p}} \qquad (1)$$

where U is the number of image authors within the test set, P_u the number of individual plants observed by the u-th author (within the test set), $r_{u,p}$ is the rank of the correct species within the ranked list of species returned by the evaluated system (for the p-th observation of the u-th author). Note that if the correct species does not appear in the returned list, its rank $r_{u,p}$ is considered as infinite. Overall, the proposed metric makes it possible to compensate the long-tail distribution effects of social data. As in any social network, few people actually produce huge quantities of data whereas a vast majority of contributors (the long tail) produce much less data. If, for instance, only one person collected an important percentage of the images, the classical mean reciprocal rank over a random set of queries is strongly influenced by the images of that contributor to the detriment of the others who only contributed with few pictures. This is a problem for several reasons: (i) the persons who produce the highest volume of data are usually the most expert ones but not the most representative of the potential users of the automatic identification tools, (ii) the large number of the images they produce makes the classification of their observations easier because they tend to follow the same protocol for all their observations (same device, same position of the plant in the images, etc.), (iii) the images they produce are also usually of better quality so that their classification is even easier.

The main conclusion we can derive from the results of Table 1 is that the classical approach to plant identification consisting of analyzing the morphology of the leaves reached its limit (we actually only observed a few slight improvements along the years). Leaf shape boundary features and shape matching techniques have actually been studied a lot during 30 years and can be considered as sufficiently mature for capturing shape information in a robust and invariant way. The limited performance is thus rather due to the intrinsic limitation of using only the leaf morphology for discriminating a large number of species. Botanists themselves are convinced that flowers and fruits are much more discriminant organs than the leaves. The phenomenon that scientists focused on leaf-based identification during years is more related to the fact that the leaf was more easy to scan and to process with state-of-the-art computer vision techniques of that period (segmentation, shape matching, registration, etc.). With the arrival of more advanced computer vision techniques, in particular the ones based on machine learning, we were progressively able to make use of other parts of the plant such as flowers or fruits. For this reason, metrics on leaf scans (or leaves on white background) were abandoned from the PlantCLEF evaluation after 2013. Only photographs of leaves in their natural environment were considered in addition to the other newly introduced views including flowers, fruits, stems and branches in their natural environment as well as photographs of the entire plant.

Table 2 gives the five-year synthesis of this new approach to plant identification that we actively promoted through PlantCLEF. The most interesting conclusion we can derive from it is that we observed considerable improvements of the scores along the years whereas the difficulty of the task was increasing as well. Actually, the number of classes almost doubled every year between 2011 and 2015, starting from 71 species in 2011 and reaching 1000 species in 2015. The increase of the performance can be explained by two major technological breakthroughs. The first one was the use of *aggregation-based* or *coding-based* image representation methods such as the Fisher Vector representation [32], that was used by the best performing system of Nakayama *et al.* [29] in 2013 and Chen *et al.* [8] in 2014. These methods consist of producing high-dimensional representations of the images by aggregating previously extracted sets of hand-crafted local features into a global vector representation. They rely on a two step process: (i) the learning of a set of latent variables that explain well the distribution of the local features in the training set (denoted as the codebook or vocabulary), and (ii) the encoding of the relationship between the local features of a given image and the latent variables. Overall, this allows to embed the fine-grained visual content of each image into a single representation space in which classes are easily separable even with linear classifiers (such as support vector machines).

The second technological step explaining the last increase of performance is without much surprise the use of deep learning methods, in particular convolutional neural networks (CNN) such as GoogLeNet [38]. In 2015, the 10 best evaluated systems were actually based on CNN. The difference of performance between them is mainly due to particular system design improvements such as the use of bagging in the best run of Choi [9]. Although, deep learning is an old research direction that has been widely studied since the end of the 80's, CNNs recently received a high

Table 2: **Five-year synthesis of the PlantCLEF challenge (plants in their natural environment)**

Year	#Species	#Images	#Evaluated systems	Score of best system	Brief description of best system
2011	71	1,469	20	0.251	. Model-driven segmentation Shape features . Random forests
2012	126	2,216	30	0.320	. Multi-scale local (color) texture SIFT + Sparse coding Spatial Pyramidal Matching . Linear SVM
2013	250	11,046	33	0.393	. Dense-SIFT, C-SIFT, Opponent SIFT HSV-SIF, self-similarity SSIM . Fisher Vectors . Linear Logistic Regression . Late fusion
2014	500	60,962	28	0.471	. ROI segmentation dense-SIFT + Color Moment . Fisher Vectors . SVM on FVs
2015	1000	113,2051	18	0.667	. GoogLeNet CNN . 5-fold bagging + Borda fusion

amount of attention caused by the impressive performance they achieved in the ImageNet classification task [26]. The force of these technologies relies on their ability to learn discriminant visual features directly from the raw pixels of the images without falling in the trap of the curse of dimensionality. This is achieved by stacking multiple *convolutional layers*, *i.e.* the core building blocks of a CNN. A convolutional layer basically takes images as input and produces as output *feature maps* corresponding to different convolution kernels, *i.e* looking for different visual patterns. Looking at the impressive results achieved by CNN's in the 2015 edition of PlantCLEF there is absolutely no doubt that they are able to capture discriminant visual patterns of the plants in a much more effective way than previously engineered visual features. Interestingly, the performance increase of CNN was not observed within PlantCLEF 2014 where one of the team was already using a CNN. The main reason is that the use of external training data was not authorized before the 2015 edition and that it is well known that CNNs require very large amounts of visual training data to be effective. The method consisting of fine-tuning a CNN pre-trained on ImageNet (such as GoogLeNet) was thus not possible before 2015 whereas it is one of the main strength of these technologies. More generally, the transfer learning capacities of CNN's are a key element for domain-specific classification problems such as plant identification for which the training set is highly imbalanced and includes many classes with few instances.

Besides purely visual concerns, we present in Table 3 the results obtained by the participants who attempted to use of the metadata associated to each image, specifically the one related to geography and seasonality. One can first see that among the large number of teams involved during the five years of the challenge, only few of them actually used the geo-location and date information. And, as a matter of fact, none of them obtained the best performance (which means that the best identification methods were always based on visual content only). Furthermore, one can see that among the measurable attempts of use of metadata, none of them got a strong improvement. The best improvement was achieved by Inria team in 2013. It was obtained by post-filtering the list of candidate species based on a flowering period histogram of each species constructed from the training set (at the week level). This difficulty of successfully using geography and seasonality is, at a first glance, quite surprising. It is actually obvious that the habitat of a given species is highly correlated with its ecological profile (for instance, we expect that plants adapted to high elevation ecosystems will not be found in coastal areas). Several reasons explain this paradox. The first one is that the occurrence data of the training set is too sparse to accurately model the distribution of the species. In the 2015-th dataset (the largest one), the average number of geo-localized observations per species was actually about 20. This is clearly insufficient to hope modelling the spatial distribution of the species on the whole territory (in some mountainous or diverse regions, environmental conditions can for instance be very different at few kilometers of distance). The second reason is that the used machine learning techniques were too straightforward to well address the problem. As discussed in section 4, species distribution modeling from occurrence data is actually still a hard problem in ecology, in particular in the context of uncontrolled observations such as the one used in the PlantCLEF challenge.

3. PL@NTNET IMPACT STUDY

As discussed earlier, Pl@ntNet is among the most advanced infrastructures in the world making use of automated identification tools for monitoring biodiversity. To measure the impact of that initiative, we did survey by email a large panel of authenticated Pl@ntnet users, *i.e.* users who created a user profile on the Pl@ntNet apps and for whom we add an email address. Each user was asked to fill an online form with the announced objective to better understand their usage of that technology and improve its functionality.

Table 3: Impact of the use of geography and seasonality for plant species identification

Year	Teams	Metadata type	Run type	Score	Improvement
2011	UAIC	GPS, Date, Author Id	. Visual . Visual + metadata	. 0.156 . 0.1	-0.056
2012	IFSC USP [5]	GPS	. Visual + metadata	. 0.16	Unknown
2012	BTU DBIS [3]	GPS	. Visual . Visual + metadata	. 0.21 . 0.2	-0.01
2013	SCG USP	GPS	. Textual	. 0.025	Unknown
2013	LIRIS [7]	GPS	. Visual + metadata	. 0.092	Unknown
2013	UAIC [34]	GPS, Author Id	. Visual + metadata	. 0.127	Unknown
2013	SABANCI-OKAN [48]	Date	. Visual + metadata	. 0.181	Unknown
2013	Inria [1]	Date	. Visual . Visual + metadata	. 0.353 . 0.385	+0.032
2014	SABANCI-OKAN [47]	Date	. Visual + metadata	. 0.127	Unknown

Among the 20 859 users, and 20 003 successfully sent emails, we received a total of 719 responses within 2 weeks. The rate of non-response to the questionnaire was thus about 96.6 percent, which is not surprising for an email survey. To get information about the reasons of non-response, we sent a second email to 5,000 of the non-respondents. We got a return of 327 people. The two main reasons for non-response were that (i) people did not (anymore) use Pl@ntNet (38 %), (ii) people used Pl@ntNet but did not see our message beforehand (38 %). The survey itself included a first part, common to all users, that was dedicated to the collect of personal information (place of residence, age, email and usage frequency of Pl@ntNet). The second part was specific to the two main types of use of the application: *professional* vs. *recreational*. It was important to create these two paths since some questions were very specific to one type of use, and we did not want to extend the questionnaire with irrelevant questions for a particular type of user. The part related to *professional* included 22 questions, such as their job, in which sector are they working (private or public), how often they use the application, to what extent this app has allowed to improve their botanical skills, etc. The part related to *recreational* included 22 questions too, such as the description of the situation that made them download the app (curiosity, gardening, hiking, etc.), if the application has changed their practices, their attention to nature, what is their interest for new functions. The survey was completed by several focus groups and interviews organised with representatives of different domains: (i) scientific domain (ecology, computer science) and citizen science, (ii) agriculture, (iii) biodiversity management, (iv) education. Our goal in this paper is not to exhaustively analyse all the results of this study (this will be done later) but to publish first conclusions with regard to the potential economic and scholar impact of Pl@ntNet.

The vast majority of respondents in the survey were located in France (85.7%) and the rest was mostly divided between Belgium (4.9%), Switzerland, Spain (1.8%) and North America (Canada 1% USA and 0.8%). This is at a first glance surprising as only 30% of Pl@ntNet users are actually in France (662,295 vs. 1,490,646 outside of France). The most likely reason of this bias is that the sent email was written in French and English only, and that the English translation was provided below the French one. Table 4 presents the age distribution of respondents. It appears that more than 65% of them are over 40 years old. We can be surprised by the small number of young respondents, as this population is usually more attracted by mobile technologies than older people. This illustrates the fact that, even if the transfer of knowledge to young people on mobile devices seems to be facilitated by the use of this device, a greater effort is needed if we want to enlarge their attractiveness to this type of initiative.

Table 4: Age of respondents

Age	Workforce	Percentage
Less than 18 y.	9	1.3%
Between 19 and 25 y.	52	7.3%
Between 26 and 40 y.	190	26.8%
Between 41 and 60 y.	276	38.9%
More than 60 y.	182	25.7%
Total	709	100%

A vast majority of users exploit Pl@ntNet for their recreation (88 %). This can explain the peaks of use noticed since 2013, during weekends. Most of the users in this category used Pl@ntNet in a garden or during a trekking. The horticultural and trekking domains are probably the two most important in which this kind of application can have a strong impact. Gardens are becoming more urban with the recent evolution of our societies. This recreational activity is motivated by a variety of factors from a stronger immersion in nature to gastronomy. The gardening market, with products that are more convenient and accessible, is then growing in popularity. Smart phones and social media start to play an important role, particularly among younger gardeners. In France, in 2015 alone, the gardening market represented 8.1 billion euros (source: promojardin[9]). The most important part of this market (1.6 billion euros) is dedicated to the acquisition of outdoor plants. The selection of the correct species based on the needs of the consumer (ornamental aspect, fruit production, shadow, speed of growth, cost of maintenance, etc.) is then a key step in gardening practices. Plant identification in the garden context can then generate a strong economic impact, in facilitating plant selection and

[9] http://www.promojardin.com/

acquisition of appropriate plant products for their management.

Based on this survey, the proportion of Pl@ntNet use for professional purposes is about 12% (which represents a volume of 1,200,000 sessions mobilized for professional activities considering the total number of sessions of over 10M). Table 5 provides the list of categories in which Pl@ntNet is used for professional activities. The most frequently represented category is landscape management (34.6%). It includes landscape workers, managers and architects, as well as foresters. The second category is more concerned with the production and/or transfer of knowledge (23.5%), that is to say, teachers (in botany, biology, horticulture), students (in horticultural production for example), trainers (landscape management, aromatherapy, herbal medicine, etc.), facilitators (botanists, nature guides) and scientists (biologists mainly). The category of ground workers represents 16% of professional respondents. This category includes farmers, nurserymen, horticulturists and gardeners.

To further illustrate the potential future impact of Pl@ntNet, Figure 1 provides a cartography of the number of identification sessions performed through the Pl@ntNet Android version in April 2016. For the countries accounting for the most users, we provide the number of sessions as well as its increase in percentage compared to the same period in 2015 (so as to illustrate the dynamic). It first shows a strong increase in the countries neighboring France (in Italy and Spain the number of sessions was actually multiplied by respectively 35 and 22). This is not surprising since there is a high intersection between the floras of those countries and the one of France, which was the starting point in the first release of the Android application in March 2014. We thus observe a geographic diffusion of the usage of the application that is related to the increasing coverage of the related species in the database as well as to the media coverage. Besides, we also observe a very strong progression in South America that is related to the release of a version of Pl@ntNet working on the Guyana flora in October 2015. Here again, we observed the same geographic diffusion phenomenon. Finally, we can also observe a relatively lower but still strong increase in North America whereas no specific version of Pl@ntNet was released there. Several factors might explain this increase including curiosity of people for such new technology or the fact that the climate of some US regions is very similar the one of Europe (so that their flora has a consistent intersection at the species level and strong intersection at the genus and family levels).

Whatever the future of the Pl@ntNet initiative in itself is, this impact study clearly shows that domain-specific mobile search technologies are attracting strong interest in our society. We can thus hypothesize that such new practices of questioning our environment will bring a lasting production of plant and animal observations. The implicit biodiversity monitoring scenario introduced in this paper is thus realistic from a societal point of view. Now it still raises brave new research challenges that will be introduced hereafter.

4. FROM EXPLICIT TO IMPLICIT PLANT BIODIVERSITY MONITORING

Whereas previous approaches to monitor plant biodiversity were based on the explicit sharing of plant observations (be they partially automated or not), the new concept we

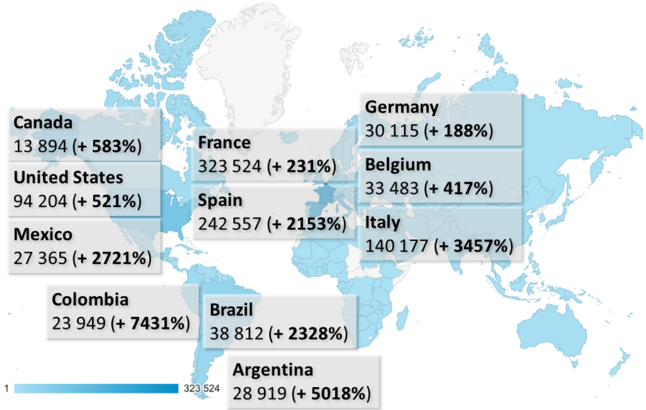

Figure 1: Cartography of the number of Pl@ntNet Android sessions in April 2016 (increase over April 2015 in parenthesis).

Table 5: Business Categories

Business Category	Workforce	%
Land Management / Environment	28	34.6
Education / Training / Research	19	23.5
Council	5	6.2
Agriculture (ground work only)	13	16
Medicine / pharmaceuticals	6	7.4
Management team	4	4.9
Other	6	7.4
Total	81	100

introduce in this paper is the implicit detection of plant occurrences in *mobile search logs* (or more generally in any stream of geo-localized user generated pictures). In recent years, we actually saw the arrival of more and more *mobile search* applications such as LikeThat, Goggles or CamFind, that allow users to get information about surrounding objects by simply photographing them (thanks to supervised classification or content-based image retrieval). These applications are still far from well recognizing any domain-specific object, but on the other side their search logs capture the user's interest about the world's objects at a very large scale and high rate. They generate quantities of geo-localized visual data that are noisy but might be used to monitor our environment and enrich its visual knowledge. In this paper, we focus on the search logs of the Pl@ntNet *mobile search* application, but in essence, the challenges we discuss could apply to any other mobile search application. As a concrete illustration, Figure 2 provides a small sample of geo-localized and dated image queries that were submitted to the Pl@ntNet application. It is likely that only a small fraction of the observations might be of interest for monitoring biodiversity. Some pictures do not contain plants at all (people, indoor scenes, mushrooms, etc.). Some others do contain plants but are so noisy or cluttered that they could not be identified. A few are not really noisy but still do not contain sufficient visual evidence to discriminate the plant (e.g. the flower is not visible whereas it is a critical information for many groups of species). Finally, many pictures do

Figure 2: A sample of the geo-localized image search logs of Pl@ntNet mobile application

represent plants that are of very limited interest for monitoring biodiversity (*e.g.* vegetables, popular horticultural plotted plants, grass, etc.). Producing accurate plant observations from such noisy and open data raises brave new research challenges that we will introduce hereafter.

Challenge 1 - Dealing with novelty and uncertainty. Knowing how much automatically predicted labels can be trusted is essential for further data processing such as human validation or direct statistical analysis. A good knowledge of the uncertainty of the automatic predictions is actually required to select the most beneficial ones (for a given scenario) or to devise robust statistical inference methods. In our implicit biodiversity monitoring scenario, any automated species detection should thus be systematically associated to a confidence score in $[0, 1]$ quantifying the probability that this prediction is true, independently from the other predictions. Doing so in the context of a noisy visual data stream such as Pl@ntNet search logs is a hard problem for two main reasons: (i) the massive presence of unknown classes in the stream (because it works in an open world) and (ii), the heavily imbalanced training set (that is inevitable when dealing with biodiversity data). When launching a new country-specific instance of Pl@ntNet, the proportion of images belonging to unknown classes can for instance be very high, up to 80%. It can remain high even in the long term because of the continuous emergence of new classes. Estimating the probability of the membership to an open set of unknown classes is thus a crucial preliminary step before being able to model the ambiguity over the known classes. This is in essence a *novelty detection* problem (see e.g. [31] for a comprehensive review) but the fact that the data set is highly imbalanced increases the difficulty of the problem. Indeed, as the majority of the known classes in the long tail only contains few training samples, they are likely to be confused with the unknown classes when using classical novelty detection algorithms. To deal with this problem, we believe it is required to primarily detect the novelty at the image level, for instance by estimating the uncertainty of the visual representation of each image during the learning process. Therefore, it might be required to consider species confusion as an input data, typically by inferring it from the annotators confusion as discussed in challenge 2. The degree of novelty of a visual content item could then be estimated as the degree of uncertainty on whether it already exists in the training set.

Challenge 2 - Enriching the training set in a collaborative way. As discussed above, one of the main sources of uncertainty when trying to recognize plants in image search logs, is the lack of training data in sufficient quantity and quality. The majority of the images in the search logs do belong to either unknown classes, *i.e.* with no training samples in the training set, or to weakly supervised classes, *i.e.* with very few training samples. A straightforward solution to reduce the uncertainty of the predictions is thus to enrich the training set. Actually, recent deep learning models, such as convolutional neural networks [26], are capable of learning very effective visual features directly from the raw image pixels but to outperform the previous methods based on handcrafted visual features, they still need to be trained on rich visual data with diverse enough visual patterns and accurate class labels. Such ideal content is unfortunately missing for the vast majority of plant species that lie in the long tail of existing data distribution (contrary to the most common ones that are over-represented). Large domain-specific collections such as Encyclopedia of Life (EOL) archives include quantities of well structured tags across many plant groups but they are not aimed at labeling precise domain-specific elements (e.g. spine, latex, branching pattern, buds, etc.), nor at covering their diversity. On the other side, computer-vision oriented data sets such as ImageNet [10] are only focused on the most popular species of the web and are too noisy from a taxonomic perspective (mix of common, species and genus names, confusions across species, horticultural plants, hybrids, etc.).

In the end, the most beneficial way to enrich the training set (and reduce the uncertainty of the predictions) is to directly annotate a fraction of the search logs themselves. Applying state-of-the-art crowdsourcing approaches in this regard is however impossible (*e.g.* [25, 37, 40]). First, the brute-force approach consisting of a quiz across the full list of species would only be tractable for the few specialists of a given flora, thus drastically limiting the potential number of annotators. Second, the very high number of classes (*i.e.* species), makes it impossible to train a complete confusion matrix for each annotator as it would require to answer to a large number of queries (typically quadratic in the number of classes). A much more promising approach is thus to devise effective *collaborative active learning* algorithms, *i.e.* learning algorithms that actively select samples to be annotated as well as annotators in a joint objective. The main underlying assumption is that even non-specialists are capable of recognizing a few tens of species (if we teach them), so that in the end, they might collectively solve complex classification tasks with thousands of classes. As in crowdsourcing algorithms, this paradigm supposes that we can model the imperfection of the annotators typically by inferring their confusion based on the labels they provide. Additionally, it requires inventing *active training* strategies aimed at training the annotators on confusions that exist within the data. Overall, *collaborative active learning* poses several fundamental questions: (i) how to optimize the selection and assignment of the unvalidated samples? (ii) how to model the learning process of the annotators to train them effectively and complementary? (iii) how to design new machine learning algorithms and/or statistical inference methods that deal with the partial knowledge of the annotators?

Challenge 3 - Using the taxonomy to reduce uncertainty. Graph-based knowledge representations such as taxonomies or ontologies are available for many domains, in particular those with high expertise such as botany. When such a rich organization of the visual class labels exists, it is likely to facilitate the estimation and reduction of the uncertainty, even though it is incomplete in some of its branches. More precisely, it allows restricting our general problem to the case where the unknown classes occurring in an uncertain visual data stream are supposed to have at least one parent in the taxonomy of the known classes. Thanks to this relaxation, challenges 1 and 2 can be revisited in a radically different manner. We can actually now have a hierarchical representation of the uncertainty, typically through hierarchical conditional probabilities. Such hierarchical structuring of the uncertainty is likely to be very effective for breaking the complexity due to extremely large numbers of classes. The automatic prediction of the uncertainty of the unlabeled visual data might for instance benefit from the knowledge of the labels structure by using it as a way of post-checking the veracity of a given prediction a posteriori. Concerning the collaborative active learning framework, both the active training of the annotators, the task assignment and the inference methods could be revised. For instance novices should start on easy to discriminate nodes of the taxonomy whereas the most advanced contributors should tackle the leaves of the taxonomy that are the most difficult to disambiguate.

Challenge 4 - Using environmental data to reduce taxonomic uncertainty. As discussed in section 2, using occurrence information (*i.e.* the geo-location and the date of the observation) did not conduct to significant identification improvement in the past PlantCLEF evaluation campaigns because of the sparsity of the occurrence data in the training set. Thus, a first naive solution could be to use a much larger occurrence data such as the one collected through the world-scale GBIF initiative. However, even with such big data, sparsity would still be an issue, in particular for the vast majority of species lying in the long tail distribution. Actually, producing masses of occurrence data, timely and globally, is precisely the objective of the implicit biodiversity monitoring scenario proposed in this paper. So that, it is somehow a chicken-egg problem. Improving plant identification systems thanks to geography would require accurate species distribution models but, on the other side, building such models requires large amounts of occurrence data and would clearly benefit from automated identification tools. A solution to that problem might rely in the use of external environmental data such as habitat maps, climate maps, soil characteristic maps, topographical maps, etc. Such data does actually less suffer from the sparsity problem and many regions of the world are well covered with such information. Thus, it might be possible to learn the ecological profile of each species by correlating its occurrences with the environmental variables and then predict the likelihood of its presence in other regions. Several issues might however be still challenging. Humans' impact does notably alter the correlation between plant habitats and environmental variables. In cities and other highly frequented places, the presence of a species is for instance rather correlated to its usage by humans (e.g. potted plants, parks, etc.). Human equipments such as roads or railways as well as human activities such as agriculture or forestry tend to quickly and deeply modify species distribution and to fragment the habitats.

Challenge 5 - Controlling observer and detection bias in species distribution models As for any presence-only data (*i.e* where information is available concerning species presence but not species absence), Pl@ntNet search logs are subject to bias due to observers being more likely to visit and record sightings at some locations than others. Such

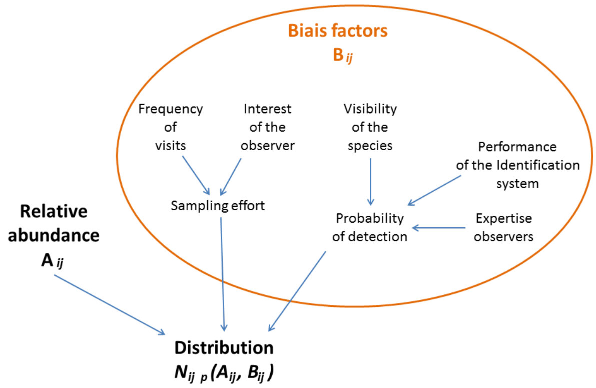

Figure 3: Probabilistic graphical model of observer and detection biais

observer bias has already been studied in some recent work on species distribution models (SDM) [42, 39]. The goal is typically to model species occurrence data through a distribution $N_{ij}\ p(A_{ij}, B_{ij})$ where Aij is the relative abundance of species i in place j (to be estimated), and B_{ij} is a more or less complex observer bias. In the context of the implicit monitoring scenario developed in this paper, modeling the bias is even more challenging. It actually depends on both *observer bias* and *detection bias* as illustrated by the probabilistic graphical model of Figure 3 that we built as a first attempt to model the problem. Incorporating taxonomic confusion in the species distribution models has in particular never been addressed before and offers brave new research perspectives at the frontier of ecological modeling and machine learning. This approach might lay the foundation to a new data-driven research field, *probabilistic taxonomy*, that has the real potential to scale up biodiversity and phenological studies to several orders of magnitude. Actually, the presence of determination errors, even with low ratios, often makes biodiversity researchers skeptical on the usefulness of crowdsourced or machine-learning based data for conducting trustable biodiversity studies. Incorporating the taxonomic uncertainty in the models and analyzing the extent to which this uncertainty yields error in SDM predictions, is thus a crucial step towards automatizing biodiversity monitoring.

5. IMPLICIT MONITORING OF INVASIVE SPECIES FROM THE PL@NTNET MOBILE SEARCH LOGS

As a first step towards evaluating the feasibility of the implicit biodiversity monitoring paradigm, we conducted an experimental study in the context of the plant task of the LifeCLEF 2016 evaluation campaign. Therefore, we created and shared a new testbed entirely composed of image search logs of the Pl@ntNet mobile application (in contrast to the previous editions of the benchmark that were based on explicitly shared and validated plant observations).

5.1 Usage scenario

As a concrete scenario, we focused on the monitoring of invasive exotic plant species. These species represent today a major economic cost to our society (estimated at nearly 12 billion euros a year in Europe) and one of the main threats to biodiversity conservation [43]. This cost can be even more important at the country level, such as in China where it is evaluated to about 15 billion US dollars annually [44], and more than 34 billion US dollars in the US [30]. The early detection of the appearance of these species, as well as the monitoring of changes in their distribution and phenology, are key elements to manage them, and reduce the cost of their management. The analysis of Pl@ntNet search logs can provide a highly valuable response to this problem because the presence of these species is highly correlated with that of humans (and thus to the density of data occurrences produced by the platform). More generally, the Pl@ntNet platform has a high potential for the monitoring and early detection of threats to biodiversity that are related to human activities.

5.2 Data

As for the training set, we used the PlantCLEF 2015 dataset enriched with the groundtruth annotations of the test images (that were kept secret during the 2015 campaign). In total, this data set contains 113,205 pictures of herb, tree and fern specimens belonging to 1,000 species (living in France and neighboring countries). Each image is associated with an xml file containing the taxonomic groundtruth (and in particular the species level ClassId), as well as other meta-data such as the type of view (fruit, flower, entire plant, etc.), the quality rating (social-based), the author name, the observation Id, the date and the geo-loc (for some of the observations).

As for the test set, we started with a randomized selection of 30K image queries that were submitted by authenticated users of the Pl@ntNet mobile application. Among this set, 3049 images had already been shared by their authors within the collaborative validation tools and were thus associated with a valid species name. The remaining pictures were distributed to three botanists in charge of manually annotating them either with a valid species name from the France flora repository or with newly created tags of their choice (and shared between them). In the period of time devoted to this process, they were able to manually annotate 4951 pictures (so as to reach 8000 images in total). Therefore, 82 new tags were created to qualify the unknown classes such as for instance *non-plant objects*, *legs* or *hands*, *UVO* (Unidentified Vegetal Object), *artificial plants*, *cactaceae*, *mushrooms*, *animals*, *food*, *vegetables* or more precise names of horticultural plants such as roses, geraniums, ficus, etc. For privacy reasons, we removed from the test set all images tagged as *people* (although they represented about 1.1% of the queries). In the end, the test set of 8,000 pictures included 3482 tagged with the newly created classes (i.e. the ones not in the training set of 1,000 species). Moreover it included 366 images belonging to a selected list of 26 potentially invasive species. This list was defined by aggregating several sources (such as the National Botanical conservatory, and the Global Invasive Species Programme) and computing the intersection with the 1000 species of the training set. At a first glance, the final number of invasive specimens in the test set might appear rather low (366). However, it represents 1.22 % of the sample of Pl@ntNet queries used to create the test set. If we confront this statistic with the millions of queries collected each year through Pl@ntNet, we could hope monitoring critical species at an unprecedented rate without any additional cost or effort for the society.

5.3 Evaluation protocol

Based on the previously described testbed, we conducted a system-oriented evaluation involving 8 different research groups who downloaded the data and ran their system. To prevent participants from tuning their algorithms on the invasive species scenario and keep our evaluation generalizable to other ones, we did not provide the list of species to be detected. Participants only knew that the targeted species were included in a larger set of 1000 species for which we provided a large training set (actually the full dataset used in PlantCLEF 2015). Participants were also aware that (i) most of the test data does not belong to the targeted list of species (ii) a large fraction does not belong to the training set of the 1000 species, and (iii) a fraction of them might not even be plants. In essence, the task to be addressed is related to what is sometimes called *open-set* or *open-world* recognition problems [2, 33], i.e. problems in which the recognition system has to be robust to unknown and never seen categories. Beyond the brute-force classification across the known classes of the training set, a big challenge is thus to automatically reject the false positive classification hits that are caused by the unknown classes (i.e. by the distractors). To measure this ability of the evaluated systems, each prediction had to be associated with a confidence score in $[0, 1]$ quantifying the probability that this prediction is true (independently from the other predictions).

5.4 Overview of the evaluated systems

The 8 participating research groups submitted a total of 29 *runs* corresponding to different configurations of their systems. 26 of them were based on CNNs and the different systems mainly differed in (i) the architecture of the used CNN, (ii) the way in which the rejection of the unknown classes was managed and (iii), various system design improvements. We give hereafter a few more details of the 3 systems that performed the best (on the invasive species). A more detailed description of these systems can be found in the working notes written by the participants and published in the CEUR-WS proceedings of CLEF 2016 (refs to be provided in final version of the paper).

Bluefield system: A VGGNet [36] based system with the addition of Spatial Pyramid Pooling, Parametric ReLU and unknown class rejection based on the minimal prediction score of training data (Run 1). Run 2 is the same as run 1 but with a slightly different rejection making use of a validation set. Run 3 and 4 are respectively the same as Run 1 and 2 but the scores of the images belonging to the same observation were summed and normalised.

Sabanci system: A CNN based system with 2 main configurations. Run 1: An ensemble of GoogleLeNet [38] and VGGNet [36] fine-tuned on both LifeCLEF 2015 data (for recognizing the targeted species) and on 70K images of the ILSCVR dataset (for rejecting unknown classes). Run 2 is the same than Run 1 but without rejection.

CMP system: A ResNet [21] based system with the use of bagging in Run 1 (3 networks) and without bagging (in Run 2).

5.5 Results

Figure 4 provides the mean Average Precision (mAP) of the best fully automated systems considering only the se-

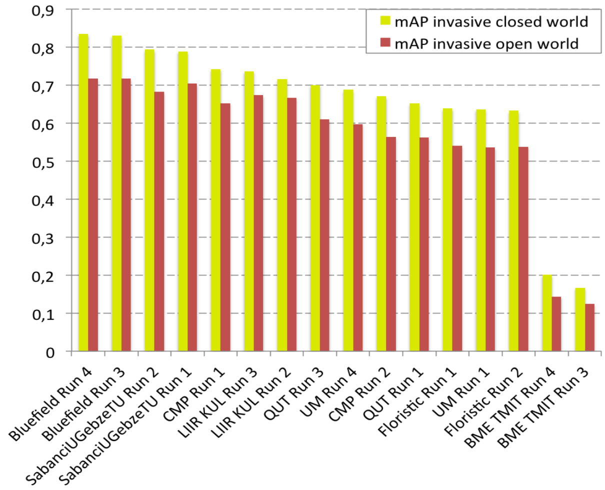

Figure 4: mean Average Precision (mAP) on the 26 invasive species in open- and closed-world

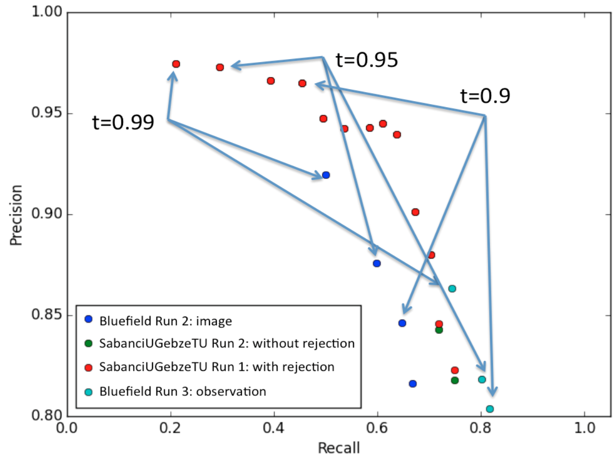

Figure 5: Precision-recall values of best systems for highly confident operating points (probability threshold $t \in [0.9, 0.99]$)

lected list of 26 invasive species as queries (only the best 2 runs of each team were kept). The mAP is computed either in open-world (*i.e.* by considering all images of the test set) or in closed-world (*i.e.* by considering only the images of the test set belonging to the 1000 species of the training set). The figure shows that the presence of the unknown classes degrades the performance of all systems in a roughly similar way. This difficulty of rejecting the unknown classes is confirmed by the very low difference between the runs of the participants who experimented their system with or without reject (*e.g.* Sabanci Run 1 vs. Run 2 or FlorisTic Run 1 vs. Run 2). On the other side, it is noticeable that all systems are quite robust to the presence of unknown classes since the drop in performance is not so high. Actually, as the CNNs were pre-trained on a large generalist data set beforehand, it is likely that they have learned a diverse enough set of visual patterns to avoid underfitting.

To better fit the implicit biodiversity monitoring scenario addressed in this paper, we completed this experiment by additional measurements more focused on high-precision operating points. If we would like the automatic predictions of the evaluated systems to be automatically integrated in an international biodiversity records database (such as GBIF), it is essential to guaranty a very high quality of the identification. Therefore, Figure 5 provides a precision/recall plot of the two best systems (in two configurations). The plot was obtained by varying the threshold of the confidence score t of each system and by measuring the recall and precision at each operating point. This experiment shows that for high precision values such as 0.99 or 0.95 only the Sabanci system evaluated in Run 1 is able to return results. However, this high precision at the price of low recall values, around 40% on average (and much lower for some of the species). In all other systems, the trust in false positives is too high and prevents reaching high precision values acceptable for biologists. This shows that the strategy of Sabanci consisting of adding a supervised rejection class is effective for managing unknown classes although it is theoretically less elegant than devising new novelty detection algorithms. Interestingly, the run of Bluefield averaging the predictions of the images belonging to the same observation provided significant improvements in recall but failed to reach high precision operating points. This does not mean that the multi-view information should not be considered as a way to deal with novelty. It rather indicates that the averaging of the predictions of the different views in not an adapted fusion scheme. Other fusion strategies should thus be explored so as to improve specificity.

6. CONCLUSIONS

The new concept we explored in this paper is the automated detection of plant occurrences in *mobile search logs* as a way to monitor biodiversity without asking the users to explicitly share and validate their observations. We showed through an impact study of the Pl@ntNet initiative that this concept is realistic from a societal point of view and that it could scale-up the world-wide collection of plant observations by several orders of magnitude. To assess the technical feasibility of such an implicit biodiversity monitoring, we summarized five years of the PlantCLEF evaluation benchmark and organized a new dedicated evaluation task within the 2016 campaign. Results show that automated plant identification systems considerably progressed during the last years thanks to successive technological advances (aggregation-based image representations and convolutional neural networks). However, in the context of very noisy content such as mobile search logs, reaching high precision is still challenging. Jointly dealing with novelty, uncertainty and highly imbalanced training data is actually a hard problem for which we suggested some new research directions. In the end, our study shows that there is still some room of improvement before being able to automatically share implicit observations within international biodiversity platforms.

7. REFERENCES

[1] V. Bakic, S. Mouine, S. Ouertani-Litayem, A. Verroust-Blondet, I. Yahiaoui, H. Goëau, and A. Joly. Inria's participation at imageclef 2013 plant identification task. In *CLEF (Online Working Notes/Labs/Workshop) 2013*, 2013.

[2] A. Bendale and T. E. Boult. Towards open world recognition. *CoRR*, abs/1412.5687, 2014.

[3] T. Böttcher, C. Schmidt, D. Zellhöfer, and I. Schmitt. Btu dbis'plant identification runs at imageclef 2012. In *CLEF (Online Working Notes/Labs/Workshop)*, 2012.

[4] D. Casanova, J. J. de Mesquita Sa Junior, and O. M. Bruno. Plant leaf identification using gabor wavelets. *International Journal of Imaging Systems and Technology*, 19(3):236–243, 2009.

[5] D. Casanova, J. B. Florindo, W. N. Gonçalves, and O. M. Bruno. Ifsc/usp at imageclef 2012: Plant identification task. In *CLEF (Online Working Notes/Labs/Workshop)*, 2012.

[6] G. Cerutti, L. Tougne, J. Mille, A. Vacavant, and D. Coquin. Understanding leaves in natural images–a model-based approach for tree species identification. *Computer Vision and Image Understanding*, 117(10):1482–1501, 2013.

[7] G. Cerutti, L. Tougne, C. Sacca, T. Joliveau, P.-O. Mazagol, D. Coquin, and A. Vacavant. Late information fusion for multi-modality plant species identification. In *Conference and Labs of the Evaluation Forum*, pages Working–Notes, 2013.

[8] Q. Chen, M. Abedini, R. Garnavi, and X. Liang. Ibm research australia at lifeclef2014: Plant identification task. In *CLEF (Working Notes)*, pages 693–704, 2014.

[9] S. Choi. Plant identification with deep convolutional neural network: Snumedinfo at lifeclef plant identification task 2015. In *Working notes of CLEF 2015 conference*, 2015.

[10] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE, 2009.

[11] K. J. Gaston and M. A. O'Neill. Automated species identification: why not? *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 359(1444):655–667, 2004.

[12] H. Goëau, P. Bonnet, J. Barbe, V. Bakic, A. Joly, J.-F. Molino, D. Barthelemy, and N. Boujemaa. Multi-organ plant identification. In *Proceedings of the 1st ACM international workshop on Multimedia analysis for ecological data*, pages 41–44. ACM, 2012.

[13] H. Goëau, P. Bonnet, and A. Joly. The lifeclef 2015 plant images identification task. In *CLEF*, Toulouse, France, 2015.

[14] H. Goëau, P. Bonnet, and A. Joly. The lifeclef plant identification task 2016. In CEUR-WS, editor, *CLEF*, CLEF2016 working notes, Evora, Portugal, 2016.

[15] H. Goëau, P. Bonnet, A. Joly, A. Affouard, V. Bakic, J. Barbe, S. Dufour, S. Selmi, I. Yahiaoui, C. Vignau, et al. Pl@ntnet mobile 2014: Android port and new features. In *Proceedings of international conference on multimedia retrieval*, page 527. ACM, 2014.

[16] H. Goëau, P. Bonnet, A. Joly, V. Bakić, J. Barbe, I. Yahiaoui, S. Selmi, J. Carré, D. Barthélémy, N. Boujemaa, et al. Pl@ tnet mobile app. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 423–424. ACM, 2013.

[17] H. Goëau, P. Bonnet, A. Joly, V. Bakic, D. Barthélémy, N. Boujemaa, and J.-F. Molino. The imageclef 2013 plant identification task. In *CLEF*, Valencia, Spain, 2013.

[18] H. Goëau, P. Bonnet, A. Joly, N. Boujemaa, D. Barthélémy, J.-F. Molino, P. Birnbaum, E. Mouysset, and M. Picard. The imageclef 2011 plant images classification task. In *CLEF 2011*, 2011.

[19] H. Goëau, P. Bonnet, A. Joly, I. Yahiaoui, D. Barthélémy, N. Boujemaa, and J.-F. Molino. The imageclef 2012 plant images identification task. In *CLEF*, Rome, Italy, 2012.

[20] H. Goëau, A. Joly, P. Bonnet, S. Selmi, J.-F. Molino, D. Barthélémy, and N. Boujemaa. The lifeclef 2014 plant images identification task. In *CLEF*, Sheffield, UK, 2014.

[21] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*, 2015.

[22] A. Joly, P. Bonnet, H. Goëau, J. Barbe, S. Selmi, J. Champ, S. Dufour-Kowalski, A. Affouard, J. Carré, J.-F. Molino, et al. A look inside the pl@ ntnet experience. *Multimedia Systems*, pages 1–16, 2015.

[23] A. Joly, H. Goëau, P. Bonnet, V. Bakić, J. Barbe, S. Selmi, I. Yahiaoui, J. Carré, E. Mouysset, J.-F. Molino, et al. Interactive plant identification based on social image data. *Ecological Informatics*, 23:22–34, 2014.

[24] A. Joly, H. Goëau, H. Glotin, C. Spampinato, P. Bonnet, W.-P. Vellinga, R. Planqué, A. Rauber, S. Palazzo, B. Fisher, et al. Lifeclef 2015: multimedia life species identification challenges. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, pages 462–483. Springer, 2015.

[25] H.-C. Kim and Z. Ghahramani. Bayesian classifier combination. In *International conference on artificial intelligence and statistics*, pages 619–627, 2012.

[26] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[27] N. Kumar, P. N. Belhumeur, A. Biswas, D. W. Jacobs, W. J. Kress, I. C. Lopez, and J. V. Soares. Leafsnap: A computer vision system for automatic plant species identification. In *Computer Vision–ECCV 2012*, pages 502–516. Springer, 2012.

[28] S. H. Lee, C. S. Chan, P. Wilkin, and P. Remagnino. Deep-plant: Plant identification with convolutional neural networks. In *Image Processing (ICIP), 2015 IEEE International Conference on*, pages 452–456. IEEE, 2015.

[29] H. Nakayama. Nlab-utokyo at imageclef 2013 plant identification task. In *CLEF (Working Notes)*, 2013.

[30] D. Pimentel, R. Zuniga, and D. Morrison. Update on the environmental and economic costs associated with alien-invasive species in the united states. *Ecological economics*, 52(3):273–288, 2005.

[31] M. A. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko. A review of novelty detection. *Signal Processing*, 99:215–249, 2014.

[32] J. Sánchez, F. Perronnin, T. Mensink, and J. Verbeek. Image classification with the fisher vector: Theory and practice. *International journal of computer vision*, 105(3):222–245, 2013.

[33] W. J. Scheirer, L. P. Jain, and T. E. Boult. Probability models for open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 36, November 2014.

[34] C. Şerban, A. Siriţeanu, C. Gheorghiu, A. Iftene, L. Alboaie, and M. Breabăn. Combining image retrieval, metadata processing and naive bayes classification at plant identification 2013. 2013.

[35] J. Silvertown, M. Harvey, R. Greenwood, M. Dodd, J. Rosewell, T. Rebelo, J. Ansine, and K. McConway. Crowdsourcing the identification of organisms: A case-study of ispot. *ZooKeys*, (480):125, 2015.

[36] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.

[37] E. Simpson, S. Roberts, I. Psorakis, and A. Smith. Dynamic bayesian combination of multiple imperfect classifiers. In *Decision Making and Imperfection*, pages 1–35. Springer, 2013.

[38] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9, 2015.

[39] J. Velásquez-Tibatá, C. H. Graham, and S. B. Munch. Using measurement error models to account for georeferencing error in species distribution models. *Ecography*, 2015.

[40] M. Venanzi, J. Guiver, G. Kazai, P. Kohli, and M. Shokouhi. Community-based bayesian aggregation models for crowdsourcing. In *Proceedings of the 23rd international conference on World wide web*, pages 155–164. ACM, 2014.

[41] E. M. Voorhees et al. The trec-8 question answering track report. In *Trec*, volume 99, pages 77–82, 1999.

[42] D. I. Warton, I. W. Renner, and D. Ramp. Model-based control of observer bias for the analysis of presence-only data in ecology. *PloS one*, 8(11):e79168, 2013.

[43] E. Weber and D. Gut. Assessing the risk of potentially invasive plant species in central europe. *Journal for Nature Conservation*, 12(3):171–179, 2004.

[44] E. Weber, S.-G. Sun, and B. Li. Invasive alien plants in china: diversity and ecological insights. *Biological Invasions*, 10(8):1411–1429, 2008.

[45] P. Wilf, S. Zhang, S. Chikkerur, S. A. Little, S. L. Wing, and T. Serre. Computer vision cracks the leaf code. *Proceedings of the National Academy of Sciences*, 113(12):3305–3310, 2016.

[46] B. Yanikoglu, E. Aptoula, and C. Tirkaz. Automatic plant identification from photographs. *Machine Vision and Applications*, 25(6):1369–1383, 2014.

[47] B. Yanikoglu, Y. Tolga, C. Tirkaz, and E. FuenCaglartes. Sabanci-okan system at lifeclef 2014 plant identification competition. In *Working notes of CLEF 2014 conference*, 2014.

[48] B. A. Yanikoglu, E. Aptoula, and S. T. Yildiran. Sabanci-okan system at imageclef 2013 plant identification competition. In *CLEF (Working Notes)*. Citeseer, 2013.