

Pl@ntNet - My Business

Alexis Joly
Inria, LIRMM
Montpellier, France

Pierre Bonnet
Cirad, AMAP
Montpellier, France

Antoine Affouard
Inria, LIRMM
Montpellier, France

Jean-Christophe Lombardo
Inria, LIRMM
Montpellier, France

Hervé Goëau
Cirad, AMAP
Montpellier, France

ABSTRACT

Pl@ntNet is a world-scale participatory platform and information system dedicated to the monitoring of plant biodiversity through image-based plant identification. Nowadays, the mobile front-end of Pl@ntNet has been downloaded by more than 4 millions users in about 170 countries and an active community of contributors produce and revise new observations everyday. This paper presents a business proposal allowing enterprises or organizations to set up their own private collaborative workflow within Pl@ntNet information system. The main added value is to allow them working on their own business object (e.g. plant disease diagnostic, deficiency measurements, railway lines maintenance, etc.) and with their own community of contributors and end-users (employees, sales representatives, clients, observers network, etc.). This business idea answers to a growing demand in agriculture and environmental economics. Actors in these domains begin to know that machine learning techniques are mature enough but the lack of training data and of efficient tools to collect them is a major breakthrough. A collaborative platform like Pl@ntNet extended with the technical innovations presented in this paper is the ideal tool to bridge this gap. It will initiate a powerful positive feedback loop boosting the production of training data while improving the work of the employees.

1 INTRODUCTION

Biodiversity informatics appeared in the 1990s as a new field aiming to conduct multi-disciplinary research at the frontier of IT and taxonomy. This research has provided shared rules and standards to facilitate implementation of exchange formats and unique resource identifiers (cf. TDWG¹). Thus, it allows analyzing big data collections produced internationally through the aggregation of data from different research institutions (e.g. Gbif, iDigBio [?] or The Plant List [4]). In addition, the recent development of participatory sciences generates new forms of production, aggregation and

exploitation of biological observations. These new initiatives, mobilizing broad networks of observers (e.g. eBird [7], iNaturalist, iSpot [6], Pl@ntNet [3], CityNature, etc.), produce massive occurrence data in a very different way than those operated in the past. The participatory sensing approach is particularly active in the birds watching field with the eBird initiative, that has become the main contributor to GBIF with over 300 million observations. Such a production is facilitated by the low number of bird species present in a given country (usually a few hundred species) allowing many people without a strong expertise for their identification to participate in such surveys.

Regarding plants, however, the large number of species on our planet (several hundreds of thousands) and the difficulty of identifying them, makes much more difficult to set up such a crowdsourcing approach. The main objective of the Pl@ntNet initiative was to fill this taxonomic gap by devising innovative machine learning and collaborative data management solutions. It was launched in 2009 by a French consortium of multimedia researchers and biologists funded by the Agropolis Foundation. Four years later, it released a first mobile application helping non experts users to identify plants thanks to content-based image retrieval techniques. In 2015, the system was upgraded with deep learning technologies allowing to learn much more effective representations and to improve the recognition performance. Nowadays, the application is becoming more and more popular especially during spring and summer. It is translated in eleven languages and has been downloaded by more than 4 millions users in about 170 countries (it is often referred to as the *shazam of plants*). An impact study [?] conducted by via usage statistics, interviews and a survey of 720 user's, concluded that the platform is reaching a strong societal and economic impact. A majority of users exploit Pl@ntNet for their recreation (88 %) and nearly 12% are running the application in a professional context. The most frequently represented professional activity 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.

This paper presents a business venture aimed at sustaining and enhancing the Pl@ntNet's infrastructure by extending it to more profitable domains than biodiversity preservation solely (while

¹<http://www.tdwg.org/>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
MM'17, October 23–27, 2017, Mountain View, CA, USA.
© 2016 ACM. ISBN 978-1-4503-4906-2/17/10...\$15.00
DOI: <https://doi.org/10.1145/3123266.3129312>

keeping it as a core objective). In a nutshell, the core idea of our business project is to facilitate the transfer of knowledge across a large variety of plant related businesses thanks to deep-learning based innovations presented in this paper. It is motivated by concrete expressions of interest issued by companies from various domains including agriculture industry, energy, transport or urban planning. In section 2, we first give an overview of the existing Pl@ntNet framework. Then, in section 3, we introduce our business project in more details. Finally, section 4 presents the technical innovations we are currently working on to boost the effectiveness of the Pl@ntNet system on a large variety of plant-related businesses.

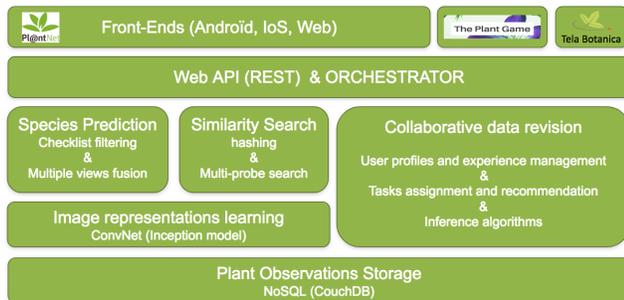


Figure 1: Pl@ntNet system architecture

2 PL@NTNET FRAMEWORK

Figure 1 presents an overview of Pl@ntNet’s architecture (at the beginning of 2017). We give hereafter some details of the main modules.

2.1 Plant Observations Storage

Plant observations are the basic data unit of Pl@ntNet information system. An observation documents evidence of a living individual plant. It includes one or several images of the plant tagged with a view type (flower, leaf, fruit, stem or entire plant) as well as some provenance data (device, author, date, etc.). Each observation is optionally associated to a geo-location and to one or more determinations, *i.e.* possible species names proposed by the author himself, and/or by other annotators, and/or by automated classifiers. A determination is tagged as valid when its confidence score (computed through different inference algorithms) exceeds some threshold. Observations are stored within a NoSQL document storage called CouchDb which is easily distributable. It is currently running on 2 servers hosted at CIRAD in France. Images are stored within a blob storage running on another server in the cloud.

2.2 Image representation learning

Image representations are computed by a convolutional neural network (CNN) that is periodically trained in a supervised manner on the observations with a valid determination name and an additional rejection class (containing non-plant pictures taken by Pl@ntNet users, *e.g.* faces, animals, manufactured objects, etc.). At the time of writing, the used CNN architecture is the inception model [8]

extended with batch normalization [?]. The network is pre-trained on ImageNet dataset [1] and periodically fine-tuned on Pl@ntNet data. The number of species (*i.e.* classes) in June 2017 was about 13 500K and the number of training images about 586K.

2.3 Species prediction

As an observation might be composed of several pictures of the observed plant, the predictions of the CNN for each searched picture need to be fused. We therefore use a weighted average of the SOFTMAX probability vectors. The weight of each picture only depends on its view type (*e.g.* the pictures tagged as *flower* are more weighted than the pictures tagged as *leaf* because flowers are much more discriminant than leaves). The values of the weights have been optimized empirically. After this fusion process, a species filtering is applied based on the *checklist* of species activated within the app (*e.g.* West Europe, North Africa, South America, etc.). By default, this checklist is automatically chosen according to the geo-location of the mobile device, but the user also has the possibility to select another one manually. Note that all the checklists share some common species. The degree of overlap depends of several parameters including the size and richness of the covered area, its environmental factors, distance between areas, etc.

2.4 Similarity Search

In addition to the most probable species, Pl@ntNet’s search engine returns the images of the dataset that are the most similar to the queried observation. This allows improving the interactive identification process by illustrating the candidate species with images consistent with the observed individual plant (the variability within the same species can indeed be high as well as the shooting protocol). This content-based image retrieval is performed through a hashing-based approximate nearest neighbors search algorithm applied on top of the 1024-dimensional feature vectors extracted by the last hidden layer of the fine-tuned CNN. The feature vector of each image is compressed into a compact binary code thanks to an unsupervised hashing method (RMMH: [?]) and its approximate k-nearest neighbors are searched by probing neighboring buckets in the hash table (using the multi-probe algorithm described in [?]). Note that a separate hash table is used for each checklist to focus the search on the right species and facilitate the distribution of the whole index. If a candidate species does not have any representative image in the returned top-k, it is not displayed to the user (actually, this adds a distance rejection criterion to the recognition process).

2.5 Front-ends and Web API

Pl@ntNet framework offers 3 main front-ends, an Android app (the most advanced and the most used one), an iOS app (being currently re-developed) and a web interface², each allowing to submit one or several pictures of a plant in order to get a list of the most likely species in return. The list of candidate species predicted for a searched plant observation is displayed in decreasing order of confidence and each species is illustrated by the most similar picture of that species. If the user selects a species, the other retrieved images are displayed by decreasing similarity scores (grouped by

²<http://identify.plantnet-project.org/>

type of view). At this stage, the user can access to richer information about the species such as all the available illustrations in the dataset (allowing to refine the identification) or descriptive web pages of the species providing useful information about the plant. Data exchanges between the search engine (server side) and the front-ends (client side) are managed through a REST-full web API using json data format and jpeg images.

2.6 Collaborative data revision and validation

One of the main strength of Pl@ntNet is that the training set enabling the automated recognition is collaboratively enriched and revised. Indeed, users can share their observation with the community (whether they have identified it or not). Several collaborative tools are then available for data validation:

- embedded revision mechanisms in the front-ends. All the contributed observations can be (re-)annotated by registered users of the Android and Web front-ends of Pl@ntNet (species name, plant organ, image quality rating, inappropriate content). The annotations are then aggregated through weighted majority voting procedures that give more weights to experimented users. The degree of experience of a user is determined according to the number of species for which he/she has contributed at least one validated observation.
- a web platform called *IdentiPlante*³ that is hosted by one of the largest network of amateur and expert botanists in the world (Tela Botanica, about 35K French speaking members). Annotations collected via this platform are aggregated in the same way than the in-app annotations.
- A gamified web application called *ThePlantGame*⁴ that is based on a set of data-driven algorithms allowing to (i) actively train annotators thanks to automatically generated quizzes, and (ii) evaluate the quality of contributors answers on new test items in order to optimize predictions (thanks to Bayesian inference and tasks assignment). This application is used as an ultimate revision tool to clean the previously validated observations that are visible in the data explorer of the application.

All these validation tools allow the app to cover a growing number of floras and species. It was initially restricted to a fraction of the European flora (in 2013) and has then been extended to the Indian ocean flora and south American one (in 2015), and last year to the north African flora (2016). A strong improvement has been done in 2017, with its extension to the North American flora, West Indies, Pacific island, and Eastern Mediterranean region.

3 BUSINESS PROPOSAL

The business proposal introduced in this paper is to allow enterprises to set up their own private collaborative workflow within Pl@ntNet information system. The main added value for them will be to work on their own business object (not only species name) and with their own community of contributors and end-users (employees, sales representatives, clients, observers network, etc.). This answers to a growing demand in agriculture, land management

and environmental economics. Actors in these domains begin to know that machine learning techniques are mature enough for their business but the lack of training data and of efficient tools to collect them is a major breakthrough. A collaborative platform like Pl@ntNet extended with the innovations presented in this paper is the ideal tool to bridge this gap. It initiates a powerful positive feedback loop boosting the production of training data while improving the work of the employees. To support this claim, we detail hereafter 4 business examples for which we've got some concrete expressions of interest (this is not an exhaustive list):

- (1) **Deficiency diagnosis in agriculture:** plants nutrient deficiency diagnosis and treatment is the core activity of many major groups in agronomy, in particular the ones specialized in plant nutrition. These groups employ thousands of employees visiting farmers all around the world in order to diagnose their cultures and sell them the most adapted nutrition product (e.g a magnesium deficiency will require product A whereas a potassium deficiency will require product B). However, the diagnosis remains a complex process requiring high expertise and/or costly laboratory analyses. An automated diagnosis in the field would therefore be a considerable gain. As discussed with some experts and companies of this domain, most deficiencies are actually associated to visual symptoms that can be directly observed on the plant. The problem is that they don't have structured training data in sufficient quality and quantity. A private Pl@ntNet project would be an ideal tool for such groups. Farm visitors employees could directly produce training data and benefit from the automated diagnosis via their mobile phone. Clients, on the other side, could themselves perform some diagnoses once the recognition is effective enough to allow non-expert using it.
- (2) **Maintenance of power lines, road, and railway equipment:** Tree pruning and weeding represents a huge cost in the maintenance of territorial facilities such as power lines, roads and railways. For instance, a large fraction of power failures are caused by branches and trees falling onto overhead power lines. Groups and companies in charge of the maintenance of these networks employ thousands of employees who regularly maintain the whole network. This process is far from optimal as it is very difficult for them to estimate the rate of growth of the different plants and the regions that should be visited in priority (the growth actually depends on several parameters including, the species, the local environmental variables, the period of the last pruning intervention). A predictive model of the stage of growth of the encountered species all over the network would therefore be a considerable gain. A private Pl@ntNet project would be an ideal tool for such groups to collect the required training data. Employees in charge of the maintenance could produce raw observations through their mobile phone (date, geo-localization and picture of the plant) while experts could annotate a part of them with stage of growth labels. A predictive model could then be trained based on the collected observations. Finally, stage

³www.tela-botanica.org/appli:identiplante

⁴www.theplantgame.com

of growth predictions could be used to plan the employee's work, globally or on the spot (to decide which plant should be pruned in priority).

- (3) **Wood pathologies diagnosis in forestry:** One of the main economical risk in temperate and tropical wood industry is the development of plant pathology at large scale, which can strongly affect wood quality and volume production. In order to avoid such damage, forest managers have to continuously prospect their forestry plots in the aim to detect potential outbreak of pathology. This detection is an expert task which is not easily transferred to new staff. A private Pl@ntNet project would allow to quickly produce, aggregate and share visual contents related to the plant pathology that occur in a forest region. It could improve the training efficiency of new employees, and save lot of money with an early pathology detection and identification process.
- (4) **Urban trees and green spaces management:** Big municipalities or companies managing urban green spaces employ thousands of people in charge of plants care and maintenance. One of the work that demands the most effort is the maintenance of trees because their dieback constitutes a risk for pedestrians and drivers. However, estimating the dieback stage is a complex task requiring high expertise (the symptoms largely depend on the species). An automated diagnosis tool would therefore help employees to decide which trees should be pruned or cut down. Here again, the lack of training data is one of the main issue and a private Pl@ntNet project would be an effective solution to collect training data while progressively improving the employees work.

4 TECHNICAL INNOVATION

To fulfill the above mentioned business opportunities and make the difference with existing approaches, the Pl@ntNet platform has to integrate a number of multimedia innovations that we are currently working on. We synthesize them hereafter and explain how they relate to the above-mentioned business scenarios:

Multi-task recognition framework with hierarchical context: the current Pl@ntNet's search engine is purely focused on species identification. A first required evolution for our business venture is thus to allow it recognize other target concepts. This could be done by re-deploying a complete new instance of the system for each new task but this would be sub-optimal and very expensive. One of the key idea of our business proposal is that the different tasks to be addressed will remain in the plant domain. Consequently, knowledge can be easily transferred from a business domain to another one and they can mutually benefit from each others. A straightforward way to achieve this is to train a multi-task neural network that learns multiple classifiers for the different tasks to be addressed (species, deficiencies, pathologies, growth stages, dieback stages, etc.). Such multi-label model allows to transfer knowledge across the tasks easily and, in fine, to reduce the amount of training data to be collected for each new business scenario. Typically, a softmax classifier is used for the species recognition whereas the

other optional labels are predicted through binary cross-entropy layers. Beyond, as most of the concepts to be recognized are species-dependent, the species prediction can be used as a context given as input to the other classifiers. This can simply be done by passing the logits of the species classifier as input to intermediate feature learning layers dedicated to the other tasks. For instance, if the visual symptom of some disease is common to a group of species, the network will learn this information to improve the recognition. In a longer term, we might explore the use of more sophisticated ways of transferring knowledge across the tasks such as domain-adversarial learning [2] or meta-learning [5].

Exif metadata as a context vector: in the business scenarios described above, some useful information about the target concepts may be missing in the visual content. The size of the observed visual patterns, in particular, can generally not be inferred from the image solely. For instance, flowers or leaves of very different sizes might appear similarly in images because of different zoom, distance to object, etc. To compensate such information loss, the solution we are considering is to use EXIF metadata as contextual information to be passed to the neural network. However, doing so is not a trivial task because EXIF data is very heterogeneous. The tags used to characterize the acquisition settings are for instance changing from a manufacturer to another, and from a camera model to another. Consequently, our primary attempt to extract structured information was not successful. The diversity of tags was too large to be manageable in the long term. The solution we are considering to solve this issue is to let the network directly learn patterns from the raw EXIF data. Therefore, we started working on an encoding scheme embedding the EXIF data into sparse vectors. The next step will be to pass this additional training data as input to the multi-task neural network. This can be done by adding a new fully connected layer on top of the raw EXIF vectors and connecting its output to each of the task-specific feature learning layer. In this way, the EXIF metadata will be used as an additional context by the classifiers.

Location-based environmental prediction: Automatically predicting the list of concepts (species, diseases, deficiencies, plant stages, etc.) that are the most likely to be observed at a given location is useful for most of the business scenarios described above. First of all, it will improve the purely content-based recognition process by reducing the list of candidate concepts that are observable at a given location. More generally, it will help organizing and optimizing the employees work by (i) improving the selection of the places to be visited in priority and (ii), helping employees in the inventory process (e.g. through a location-based recommendation service). However, learning accurate location-based predictive models in a generic way is not a trivial task because of strong sampling bias in the participatory training data. Actually, in such data, the geographical density of observations is often more correlated to the sampling effort than to the real abundance of the observed objects. Thus, in ecology, the prediction of the presence (or absence) of a given species at a given location, is usually based on environmental variables that characterize the environment encountered at that location (i.e. climatic data, topological data, occupancy data, etc.). The geo-coordinate (w_i, h_i) of an observed specimen x_i is used to

extract a d -dimensional environmental descriptor $f(w_i, h_i)$ characterizing the environment at that location, and then, the descriptors of all the specimens of a given species are passed to a regression model aimed at learning the ecological niche of that species. Inspired by such practice and by the one in the multimedia field, we started working on deep learning techniques aimed at learning more powerful environmental representations for any concepts (species, diseases, deficiencies, etc.). First of all, we propose using tensors of image patches instead of vectors as input environmental data. Combined with the use of a convolutional neural network (CNN), this allows learning complex spatial patterns characterizing the local environment of the observation in a finer way. Typically, our CNN is trained on multi-channel images of 64x64 pixels, each channel corresponding to a single environmental variable measured on a grid of 64x64 neighboring points (e.g. average, min and max temperature, soil occupation, elevation, evapotranspiration, distance to roads or water surfaces, etc.). Second, as for the visual recognition, we suggest learning a single multi-task CNN for all concepts rather than distinct models such as done in ecology. Learning a single multi-task CNN allows transferring knowledge across the different concepts of a task as well as across the different tasks (e.g. the magnesium deficiency of a species A at a given location might be correlated to the presence of a fast-growing invasive species B). Ongoing experiments of our model shows very promising results and a publication to be submitted to a special issue of *Multimedia Tools and Applications* journal is currently in preparation. This new technology will be one of the main technological asset of the business venture presented in this paper (in addition to the automated recognition and participatory validation tools). Further possible evolution includes the integration of additional geo-localized input data such as satellite imagery, co-occurrence data with other organisms, Flickr data, street-view trekker data, etc.

4.1 Front-ends and API developments

From a usage perspective, the integration of a new private business scenario within Pl@ntNet front-end has to remain as simple as possible. Once the user is authenticated, she/he will see the private projects to which she/he participates in the existing *Project* screen of the application⁵. If she/he selects one of this project, all the other screens will be contextualized accordingly (as for the public projects and micro-projects already available within the app). The *Explorer* screen will switch to the list of concepts and data related to the selected project. The *Novelty* screen will display only the contributions of the people participating to that projects. The annotation screens will be enriched by the business-specific tags. Concerning the location-based predictions, we are considering two complementary developments of the app: (i) an "around me" mode allowing to restrict the *Explorer* screen to the list of the most likely concepts observable at the user's position, and (ii) a new *Map* screen allowing to visualize the predicted spatial distribution of a given concept (typically via an interactive heat map). For enterprises who

⁵It is preferable to look to the android version for a better understanding: <https://play.google.com/store/apps/details?id=org.plantnet>. The iPhone app is older and will be upgraded in about 6 months

would like to go further in the integration of Pl@ntNet services in their own infrastructure or information system, a secured API will be made accessible to them (data access, recognition, similarity search, location-based prediction, etc.).

5 CONCLUSION

This paper presented a business venture aimed at extending the Pl@ntNet infrastructure to adapt its capacity to several different economical contexts related to plants. One of the core idea is to benefit from recent advances in deep learning such as multi-task learning, hierarchical context and transfer learning to jointly improve the performance of various plant-related tasks such as the recognition of plants nutrient deficiency, plant pathology, or plant dieback. We believe such specialization and cross-fertilization of the target business domains is a key to succeed in the emerging mobile search market. Furthermore, the collected observations will be used for learning brave new location-based prediction services that will be of a great added value for the clients work.

REFERENCES

- [1] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 248–255.
- [2] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. 2016. Domain-adversarial training of neural networks. *Journal of Machine Learning Research* 17, 59 (2016), 1–35.
- [3] Alexis Joly, Pierre Bonnet, Hervé Goëau, Julien Barbe, Souheil Selmi, Julien Champ, Samuel Dufour-Kowalski, Antoine Affouard, Jennifer Carré, Jean-François Molino, and others. 2016. A look inside the Pl@ntNet experience. *Multimedia Systems* 22, 6 (2016), 751–766.
- [4] The Plant L. 2013. Version 1.1. Published on the Internet. *The Plant List*, <http://www.theplantlist.org/> (2013), accessed 1st January.
- [5] Sachin Ravi and Hugo Larochelle. 2016. Optimization as a model for few-shot learning. (2016).
- [6] Jonathan Silvertown, Martin Harvey, Richard Greenwood, Mike Dodd, Jon Rosewell, Tony Rebelo, Janice Ansine, and Kevin McConway. 2015. Crowdsourcing the identification of organisms: A case-study of iSpot. *ZooKeys* 480 (2015), 125.
- [7] Brian L Sullivan, Jocelyn L Aycrigg, Jessie H Barry, Rick E Bonney, Nicholas Bruns, Caren B Cooper, Theo Damoulas, André A Dhondt, Tom Dieterich, Andrew Farnsworth, and others. 2014. The eBird enterprise: an integrated approach to development and application of citizen science. *Biological Conservation* 169 (2014), 31–40.
- [8] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1–9.