Towards the recognition of the world’s flora

Alexis Joly, Hervé Goëau, Valeriu Codreanu, Jean-Christophe Lombardo
Plant identification is crucial for sharing and accessing knowledge about plants
  – Food crisis
  – Biodiversity crisis

But the taxonomic gap is a tricky problem
  – Traditional tools only suitable for specialists
  – Less and less specialists

Particularly in south countries with the richest biodiversity
An innovative **citizen science** platform making use of **machine learning** to help people **identify plants** through their mobile phone.
- 11 languages
- 17K species (illustrated by 800K revised images)
- 23 projects & micro-projects (e.g. asian plants, trees of South Africa, etc.)
- 30M raw plant images
- 55M sessions / 192M screen views
- 12K followers on social networks

Last 12 months: 3,352,788 users in 235 countries

| More than 5 sessions | 1,469,423 |
| More than 10 sessions | 876,698 |
| More than 25 sessions | 172,666 |
| More than 100 sessions | 14,167 |

1. United States 412,062 (12.33%)
2. France 367,001 (11.04%)
3. Germany 191,108 (5.68%)
4. Italy 160,208 (4.83%)
5. Spain 126,640 (3.83%)
6. Brazil 110,821 (3.33%)
7. United Kingdom 77,864 (2.33%)
8. India 44,408 (1.34%)
9. Netherlands 61,830 (1.85%)
10. Canada 57,159 (1.74%)
Technology: Convolutional Neural Networks

\[ a_j^l = \sigma \left( \sum_k w_{jk}^l a_{k}^{l-1} + b_j^l \right) \]
Technology:
Convolutional Neural Networks

Species 1
Species 2
Species N

Similarity search index
(high-dimensional hashing)
Technology:
Convolutional Neural Networks

Billions of adjustable parameters (weights)
Requires high computing resources (GPUs or large clusters of CPUs)
Transfer learning (fine-tuning)

Problem: CNNs require huge training data to learn the billions of parameters
Solution: Learn domain specific features by transfer learning

1. Train CNN on a generalist image dataset with millions of images
2. Keep the weights of the lowest layers but remove/reset the top layers
3. Feed forward and back-propagate new domain specific images
Evaluation

- Pl@ntNet organizes a world-wide challenge since 2011
- Tens of research teams working on Pl@ntNet data
- **System-oriented** benchmarks/competitions
## PlantCLEF

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Espèces</td>
<td>71</td>
<td>126</td>
<td>250</td>
<td>500</td>
<td>1,000</td>
<td>1,000</td>
<td>10,000</td>
</tr>
<tr>
<td>Images</td>
<td>5,400</td>
<td>11,500</td>
<td>26,077</td>
<td>60,962</td>
<td>113,205</td>
<td>121,205</td>
<td>1.2 M</td>
</tr>
<tr>
<td>Nb. of particip.</td>
<td>8</td>
<td>11</td>
<td>12</td>
<td>22</td>
<td>15</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>Best perf.</td>
<td>0,209</td>
<td>0,38</td>
<td>0,393</td>
<td>0,456</td>
<td>0,652</td>
<td>0,742</td>
<td>0,92 !</td>
</tr>
</tbody>
</table>

### Timeline:
- **2011**: Shape descriptors
- **2012**: Bags of Words & Support Vector Machine
- **2013**: Fisher vectors
- **2015**: Arrival of deep learning
- **2016**: Convolutional neural networks
PlantCLEF 2018: Experts vs. Machines plant images identification

- 9 of the best of the best experts of the French flora
- 100 obs. including very difficult taxonomic groups
PlantCLEF vs.

Deep learning systems

- Mario TSA Berlin - Noisy
- Mario TSA Berlin - Filtered
- KDE TUT Mixed
- KDE TUT - Noisy
- CMP

Best expert
Is the problem solved? Not really...

World: 300K species

LifeCLEF: 10K species

Pl@ntNet: 17K species

Encyclopedia of Life: 50K species
Is the problem solved? Not really...

![Graph showing the relationship between number of images and accuracy. The graph indicates that 95% accuracy is achieved with a small number of images, 70% accuracy with a moderate number, and 50% accuracy with a large number.]
The Big One

- We did query Bing and Google image with 300K species names
  - Using ThePlantList: the first effort to list all plants on earth
- We collected 12 million images of 294K plant species (1.5 Tb):
  - Expert data (Encyclopedia of Life, 350K images) + Citizen science data (Pl@ntNet data, 400K images) + Web data (11 M images)
- Highly imbalanced distribution: only 50K species with more than 10 images, 50% with 1 images)
- Noise: depends on the species

“Arnica montana”
Challenges/questions

Scalability to hundreds of thousands of classes
- Which hardware?
  - Memory usage: last layer is 300 times larger than state-of-the-art models
  - To distribute or not to distribute?: communication cost, large batch size
  - CPU vs GPU?
- Which network architecture?
  - Convergence of state-of-the-art models? No guaranty
  - Do we need a new dedicated architecture?
  - Acceptable training time?
- Quality of the learned models?
  - Top-1, top-5, top-30 accuracy? On average? In the long tail?
  - Robustness to noise in the training data?
GENCI proposed us to be beta-tester of 2 prototype platforms:

- **Oussant: GPU cluster** hosted by IDRIS (IBM OpenPOWER platform)
  - 12 nodes IBM Power Systems x 4 GPU Nvidia P100 + Infiniband
  - IBM powerAI framework v4: Caffe-DLL & TensorFlow-DLL
- **Irene: CPU cluster** hosted by CEA (Intel skylakes platform)
  - 1600 nodes x 48 Intel Skylakes
  - Intel-CAFFE library
- Preparatory phase on **CINES CPU clusters** (Intel-CAFFE library)
  - **Occigen**: 3306 nodes x 2 Intel processors (12-14 cores)
  - **Frioul**: 48 nodes x Intel KNL processor (68 cores)
Evaluation methodology: set up

- Use state-of-the-art ConvNet with good size/performance tradeoff
  - Inception v2, ResNet-50
  - Extend them to 294K classes

- Distribution with synchronous Stochastic Gradient Descent
  - synchronize the gradients of N learners through a collaborative reduction/communication (such as allreduce)
  - High-dimensional = millions of values
  - Best performances in literature
    - FaceBook, ImageNet in 1 hour, 32x8 GPU P100, All reduce, batch size=8192
    - IBM powerAI, ImageNet in 50 min, 64x4 GPU P100, Multi-ring, batch size=8192
    - Intel-CAFFE, ImageNet in 28 min, 1.5K Intel KNL (100K cores), MLSL, batch size=48K
    - Preferred Networks, ImageNet in 15 min, 1K GPU P100, All reduce, batch size=32K
Evaluation methodology: test set

- **30K never published images** of expert botanists
  - Stored on their local disks
  - Complex groups in the long tail of the distribution
    - 342 Orchids species
    - 1K Guyana species
    - 469 Alpine species
    - 75 Grass species

- **PlantCLEF 2017 test set (25K images)**
  - 1K species living in America and Europe (including common ones)
  - Never published labels
Ouessant experiments (1/4)

By Hervé Goëau, data scientist Pl@ntNet (CIRAD / Inria)

- Encountered difficulties: feedback from a data scientist without experience in HPC or distributed deep learning
  - **File systems / inodes issues**: quota exceeded notifications, file creation errors, etc.
  - **No internet access**: no wget, no curl to download pre-trained models, tests, etc.
  - **Lack of documentation**
  - **Limitation of the installed frameworks**: old versions, no data augmentation, no shuffling, etc.
  - **Jobs limitation** (20h00 & 4 nodes)

- **Within the allocated time**: No efficiency gain observed in multi-nodes
Ouessant experiments (2/4)

By Hervé Goëau, data scientist Pl@ntNet (CIRAD / Inria)

- **Training models “from scratch” was not possible**
  - Several weeks on a single node
  - Without guaranty of (good) convergence
  - With a 20h jobs limitation

- **Succeeded in training a model at the scale of the world’s flora using transfer learning**
  - Inception v2 pre-trained on ImageNet and fine-tuned on 294K species in 2 steps
    1. Freeze the model except the last layer
    2. Fine-tune all layers
  - About **60h of training** on 1 node with 4 P100 GPUs
Ouessant experiments (3/4)

By Hervé Goëau, data scientist Pl@ntNet (CIRAD / Inria)

- The model works! state-of-the-art performance on PlantCLEF 2017 dataset (without using ensembles)

Our world’s flora model (with different testing configurations: data augmentation, post-filtering, duplicates removal, multi-image)
Performance in the long tail is low but fair with regard to 294K classes

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Top1 accuracy (single image)</th>
<th>Top1 accuracy + multi-image</th>
<th>Top5 accuracy + multi-image</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orchids</td>
<td>0.04</td>
<td>0.12</td>
<td>0.22</td>
</tr>
<tr>
<td>Alpine</td>
<td>0.19</td>
<td>0.25</td>
<td>0.40</td>
</tr>
<tr>
<td>Guyana</td>
<td>0.07</td>
<td>0.07</td>
<td>0.12</td>
</tr>
<tr>
<td>Grasses</td>
<td>0.37</td>
<td>0.57</td>
<td>0.71</td>
</tr>
<tr>
<td>Random</td>
<td>0.000003</td>
<td>0.000003</td>
<td>0.000015</td>
</tr>
</tbody>
</table>
CINES experiments (1/5)

- Team
  - Valeriu Codreanu & Damian Podareanu (Research engineers at SurfSara, state-of-the-art results on 1K Intel Skylake)
  - Jean-Christophe Lombardo (Research engineer at Inria - Pl@ntNet)
  - Gabriel Hautreux (HPC engineer, CINES/GENCI)
  - Vikram A Saletore (Principal Engineer for Artificial Intelligence Products at Intel)

- Objective
  - Scaling Deep learning on CPUs using INTEL-CAFFE (optimized for skylakes CPUs)
  - CEA Irene cluster (1600x48 Skylake hearts) in July (machine delivery)

- Preparatory phase on Occigen & Frioul CPU cluster from CINES
  - Occigen: 3306 nodes x 2 Intel processors (12-14 cores)
  - Frioul: 48 nodes x Intel KNL processor (68 cores)
Encountered difficulties

- Intel-CAFFE (MLSL library) requires a password less ssh connexion for initialization (only possible to run in interactive mode)
- Protobuf library is limited to 2Gb files: impossible to serialize ResNet-50 model with 275K classes → dimensionality reduction trick

```
ResNet-50
294K classes
Network size: 2.3GB
```

```
ResNet-50
294K classes
Network size: 1.8GB
```
## Scaling efficiency experiments

- Lustre striping makes a big difference

<table>
<thead>
<tr>
<th></th>
<th>2 nodes</th>
<th>32 nodes</th>
<th>Scaling efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>No striping</td>
<td>47.5 img/s</td>
<td>303 img/s</td>
<td>40.1%</td>
</tr>
<tr>
<td></td>
<td>23.7 img/s/node</td>
<td>9.5 img/s/node</td>
<td></td>
</tr>
<tr>
<td>Lustre striping</td>
<td>47.5 img/s</td>
<td>688 img/s</td>
<td>90.7%</td>
</tr>
<tr>
<td>stripe count: 64</td>
<td>23.7 img/s/node</td>
<td>21.5 img/s/node</td>
<td></td>
</tr>
<tr>
<td>stripe size: 32M</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Scaling efficiency experiments

Broadwell (BDW28) scaling results:
2 nodes; global batch size: 128; Throughput: 28.45 img/s/node. Aggregate throughput: 56.9 img/s
32 nodes; global batch size: 2048; Throughput: 25.6 img/s/node. Aggregate throughput: 819 img/s
64 nodes; global batch size: 4096; Throughput: 25.1 img/s/node. Aggregate throughput: 1606 img/s
128 nodes; global batch size: 8192; Throughput: 24.6 img/s/node. **Aggregate throughput: 3150 img/s**
→ 86.5% scaling efficiency when going from 2 to 128 BDW nodes

Haswell (HSW24) scaling results:
128 nodes; global batch size: 8192; Throughput: 20.15 img/s/node. Aggregate throughput: 2580 img/s
→ 82.2% scaling efficiency when going from 2 to 128 HSW nodes
Succeeded to learn two new models on Frioul and Occigen CPU clusters

<table>
<thead>
<tr>
<th></th>
<th>Top1 accuracy (all world flora test sets)</th>
<th>Top5 accuracy (all world flora test sets)</th>
<th>Training time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ouessant: 1 node - 4 x P100 Inception v2 fine-tuned 10 epochs</td>
<td>0.336</td>
<td>0.437</td>
<td>60 hours 6 hours/epoch</td>
</tr>
<tr>
<td>Frioul: 40 KNL nodes ResNet-50 fine-tuned 37 epochs</td>
<td>0.355</td>
<td>0.440</td>
<td>37 hours 1 hour/epoch</td>
</tr>
<tr>
<td>Occigen: 128 nodes (BDW) ResNet-50 from scratch 100 epochs</td>
<td>0.363</td>
<td>0.449</td>
<td>28 hours 17 minutes/epoch</td>
</tr>
</tbody>
</table>
Conclusion/perspectives

Conclusions
- State-of-the-art CNNs scale to 300K classes (without much modifications)
- Synchronous SGD on hundreds of CPU nodes provides high scaling efficiency but this requires significant know-how
- Training data remains a core problem

Perspectives
- Irene cluster: 1600 skylake nodes
- Inria Project Lab HPC-Deep Learning: PhD on the joint optimization of network architecture and resource allocation
Thank you