

# Identification d'espèces avec réseaux de neurones profonds

Hervé Goëau  
Mini-séminaire AMAP "réseau neuronal profond" 20 octobre 2017

# Retour d'expérience

Exploitation des techniques de deep learning appliquées à :

Identification d'espèces de plantes

*photographies "terrain"*



*scans d'herbiers*



Annotation semi-automatique  
par type de vues



Problématiques de passage à l'échelle



Identification d'espèces  
d'oiseaux basée sur le son



# Retour d'expérience

## Le coût en temps

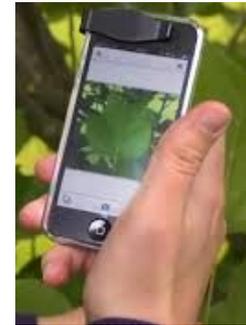
*code, préparation des données*



*pour l'apprentissage*



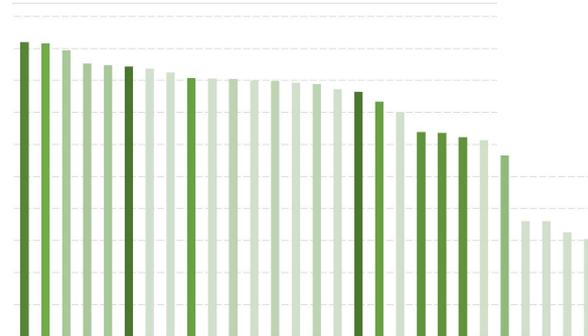
*déploiement en production*



**Le coût financier.....**



**.....et le ratio performance / coût**

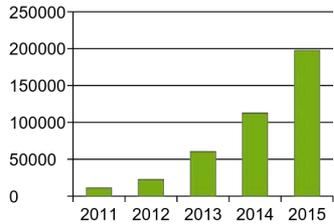




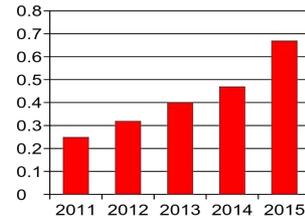
# Identification de plante par l'image

Le cercle vertueux fonctionne !

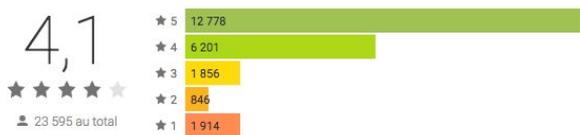
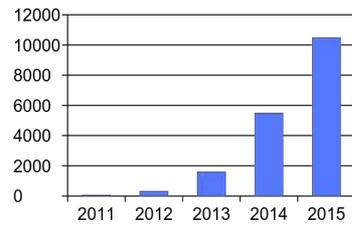
# validated images



Identification performances



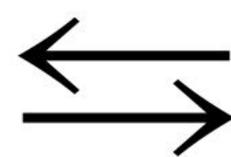
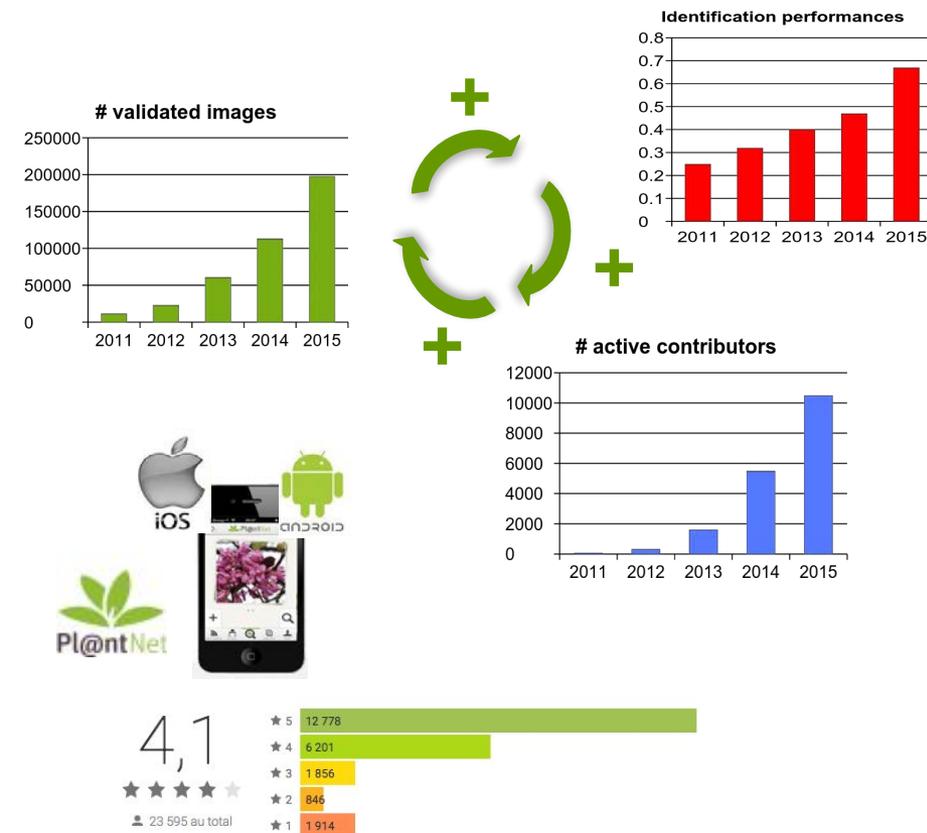
# active contributors



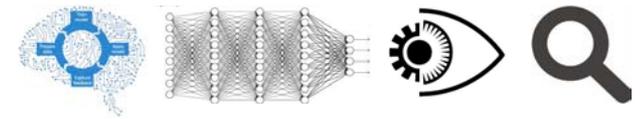
# Identification de plante par l'image

Le cercle vertueux fonctionne !

Evaluations



Recherche en informatique



international et annuel  
(conférence CLEF)



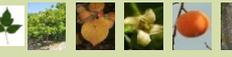
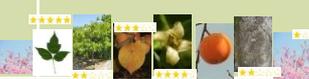
tâche de classification :

- *train (+validation)*
- *test*

# 7 années de progrès



- . De scans de feuilles produits par des experts à des observations multi-images "terrain" par des milliers d'observateurs
- . De quelques dizaines d'espèces d'arbre au passage à l'échelle d'une flore entière

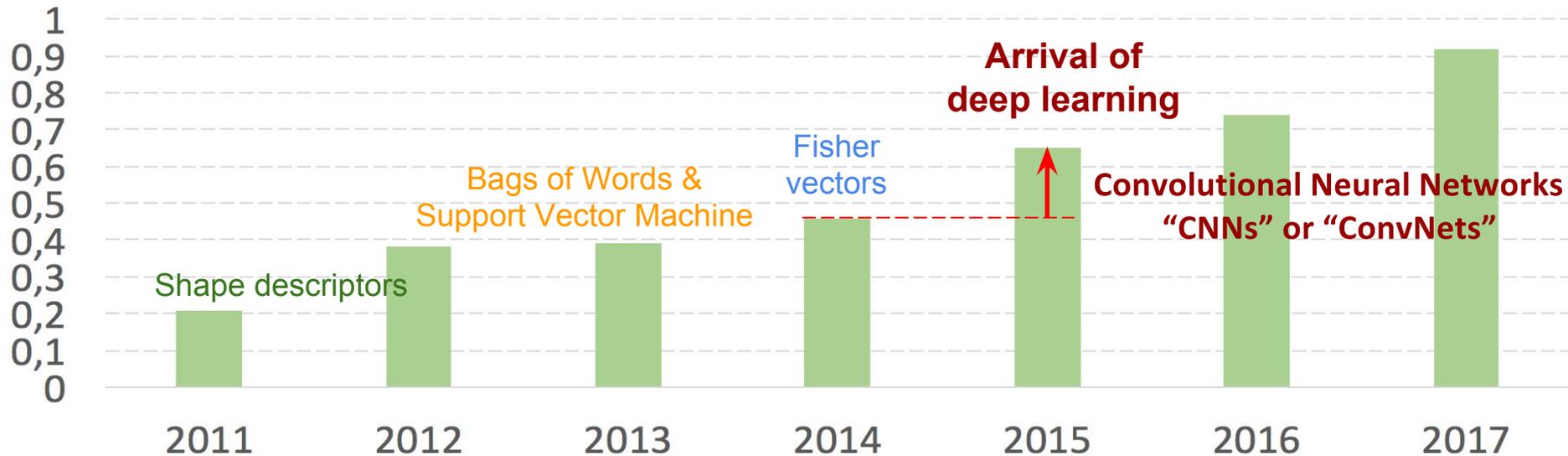
	2011	2012	2013	2014	2015	2016	2017
Espèces	71	126	250	500	1,000	1,000	10,000
Images	5,400	11,500	26,077	60,962	113,205	121,205	1,2 M
Type							
Part	8	11	12	22	15	16	17
Perf. (MMR)	0,209	0,38	0,393	0,456	<b>0,667</b>	<b>0,742</b>	<b>0,92</b>

# 7 années de progrès



. De scans de feuilles produits par des experts à des observations multi-images "terrain" par des milliers d'observateurs  
 . De quelques dizaines d'espèces d'arbre au passage à l'échelle d'une flore entière

	2011	2012	2013	2014	2015	2016	2017
Espèces	71	126	250	500	1,000	1,000	10,000
Images	5,400	11,500	26,077	60,962	113,205	121,205	1,2 M
Type							
Part	8	11	12	22	15	16	17
Perf. (MMR)	0,209	0,38	0,393	0,456	<b>0,667</b>	<b>0,742</b>	<b>0,92</b>



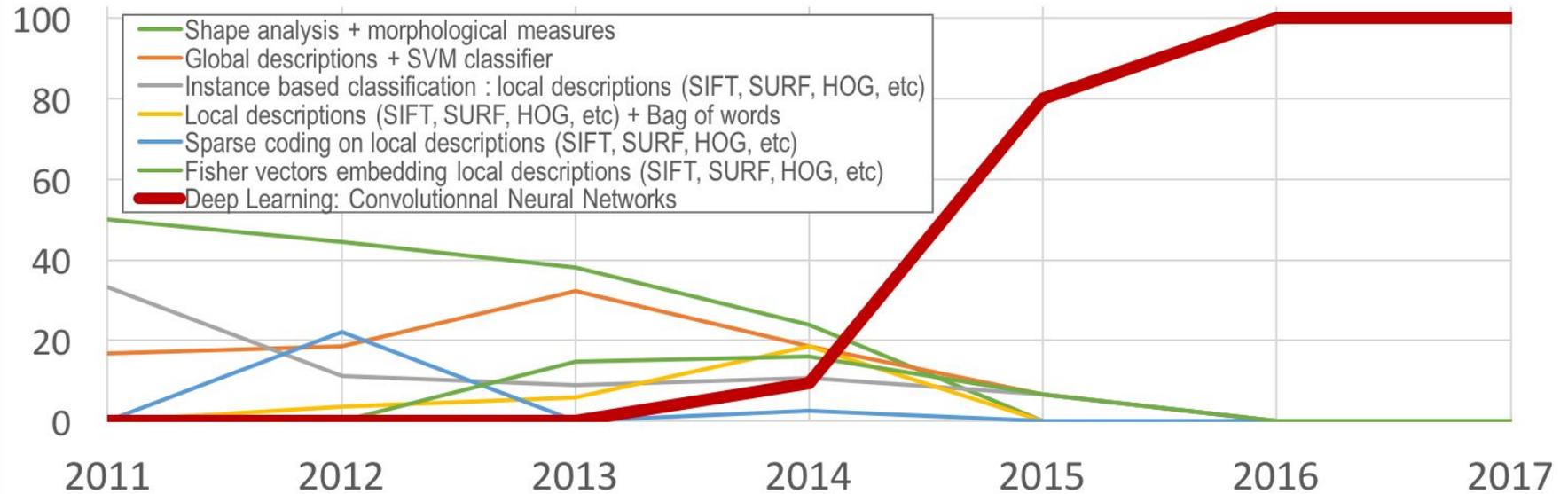
# 7 années de progrès



. De scans de feuilles produits par des experts à des observations multi-images "terrain" par des milliers d'observateurs  
 . De quelques dizaines d'espèces d'arbre au passage à l'échelle d'une flore entière

	2011	2012	2013	2014	2015	2016	2017
Espèces	71	126	250	500	1,000	1,000	10,000
Images	5,400	11,500	26,077	60,962	113,205	121,205	1,2 M
Type							
Part	8	11	12	22	15	16	17
Perf. (MMR)	0,209	0,38	0,393	0,456	<b>0,667</b>	<b>0,742</b>	<b>0,92</b>

## Popularité des méthodes (% de):

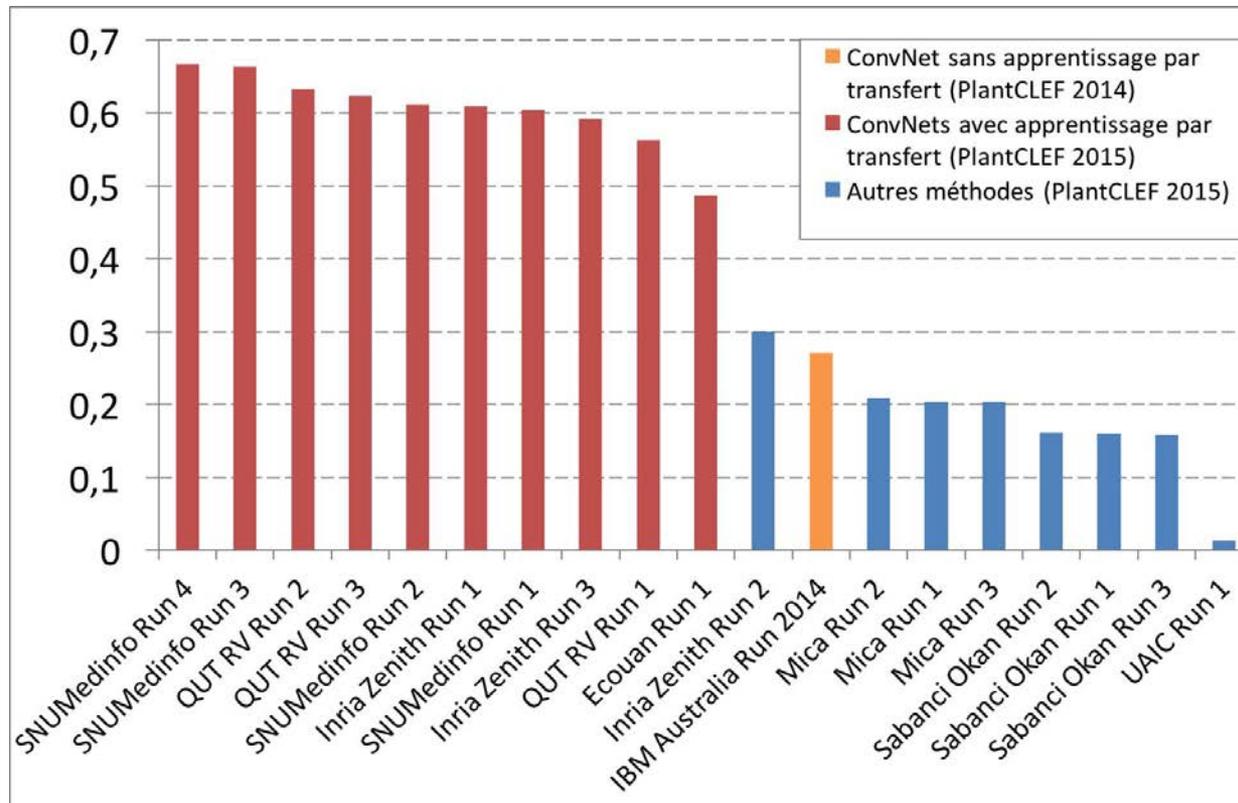


# 7 années de progrès

## La puissance du “transfer learning”

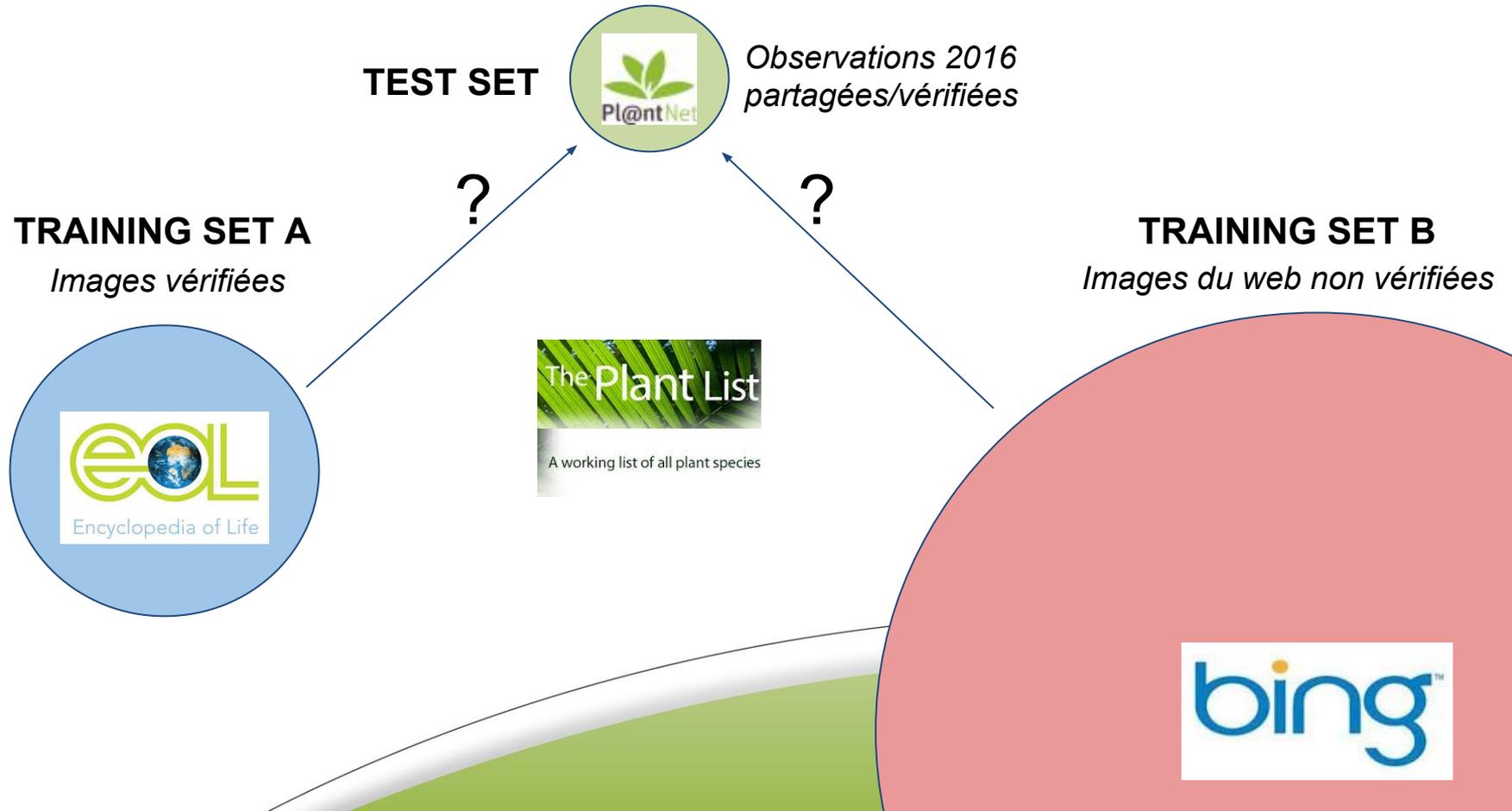


	Trademark Logos	Car models	Paris Buidlings	Aircraft models	Bird species	Flower species
GoogLeNet trained from scratch	67.7%	59.3%	55.3%	72.7%	24.4%	59.5%
GoogLeNet pre-trained on ImageNet	<b>87.5%</b>	<b>79.9%</b>	<b>71.3%</b>	<b>88.1%</b>	<b>72.4%</b>	<b>89.5%</b>



# Plant Identification Based on Noisy Web Data: the Amazing Performance of Deep Learning (PlantCLEF 2017)

- 10 000 espèces : comparable à l'échelle d'une flore semi-continentale
- intérêts d'intégrer des données externes potentiellement bruitées ?
- asymétrie train / test



# Test set



The Plant List

Pl@ntNet  
shared &  
validated  
observations  
during 2015/16



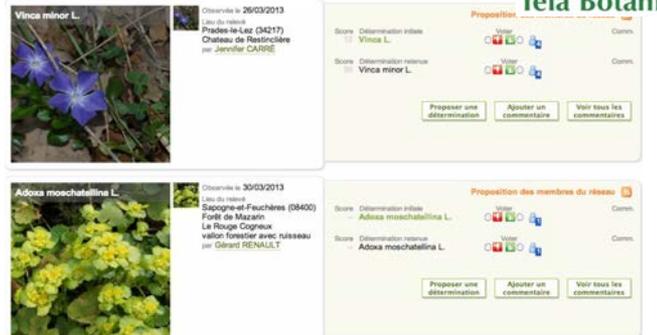
12k species  
285k observations  
430k images  
100k users

Matched species names  
with the world plant list



4,5k species  
130k observations  
169k images  
47k users

Positive votes on species  
(collaborative validation)



1125 species  
17 886 observations  
25 190 images  
9691 users



QUALITY



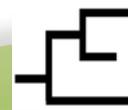
AUTHOR



DATE



LOCALIZATION



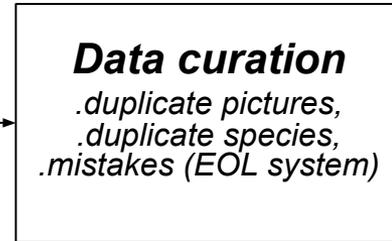
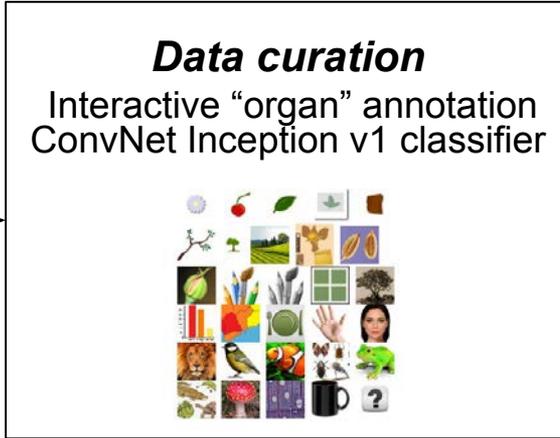
TAXONOMY



ORGAN TAG

# Training dataset (trusted)

815k pictures  
233k species



**256,287 pictures**  
**10,000 species**



About 250k herbarium sheets removed!

# Training dataset (trusted)

## “Coûts” du classifieur organe / vue

- **code, préparation des données :**
  - 5 jours (mixe une petite partie ImageNet + datastore PI@ntNet)
- **apprentissage :**
  - ConvNet de base PI@ntNet : GoogleNet modifié (batch normalization + PReLU)
    - préentraîné 15 jours d'apprentissage sur GeForce 4Go, 600 €
  - Fine-tuning sur 29 classes, 59 000 images
    - 90 minutes sur Quadro M4000 8Go, 600 € + 2400 € pour la station de travail
- **déploiement et annotations :**
  - 3 GPUs en parallèle sur 2 stations de travail :
    - 180 images par minute
    - 3 jours pour pré-annoter toute la base train EOL
  - Validation / annotation par lot : 8 000 images / heure

### **Data curation**

Interactive “organ” annotation  
ConvNet Inception v1 classifier



# Training dataset (noisy)



Extends the trusted dataset  
up to **150** additional images per species  
No data curation

**1.1M pictures**  
**10k species**

*“Arnica montana” -site:eol.org*



*2 mois*

# Training dataset (noisy)



## “Non plant”



## “Irrelevant” plant



## “Taxonomic” noise

*Geranium  
brevicaule*



*Rubus canadensis*  
*Rubus nivalis*  
*Rubus pensilvanicus*  
*Rubus pedemontanus*  
*Rubus deliciosus*  
*Rubus polyanthemus*



# Datasets & Metric



	Observations	Images	Species
 Test	<b>18k</b>	25k	1.1k
 Training Noisy	? (1.1M)	<b>1.1M</b>	10k
 Training Trusted (EOL) <small>Encyclopedia of Life</small>	? (256k)	<b>256k</b>	10k
 Training Trusted (PlantCLEF 2016)	< 42k	< 115k	<1k (880)

Mean Reciprocal Rank

$$\text{MRR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i}$$

# Participation

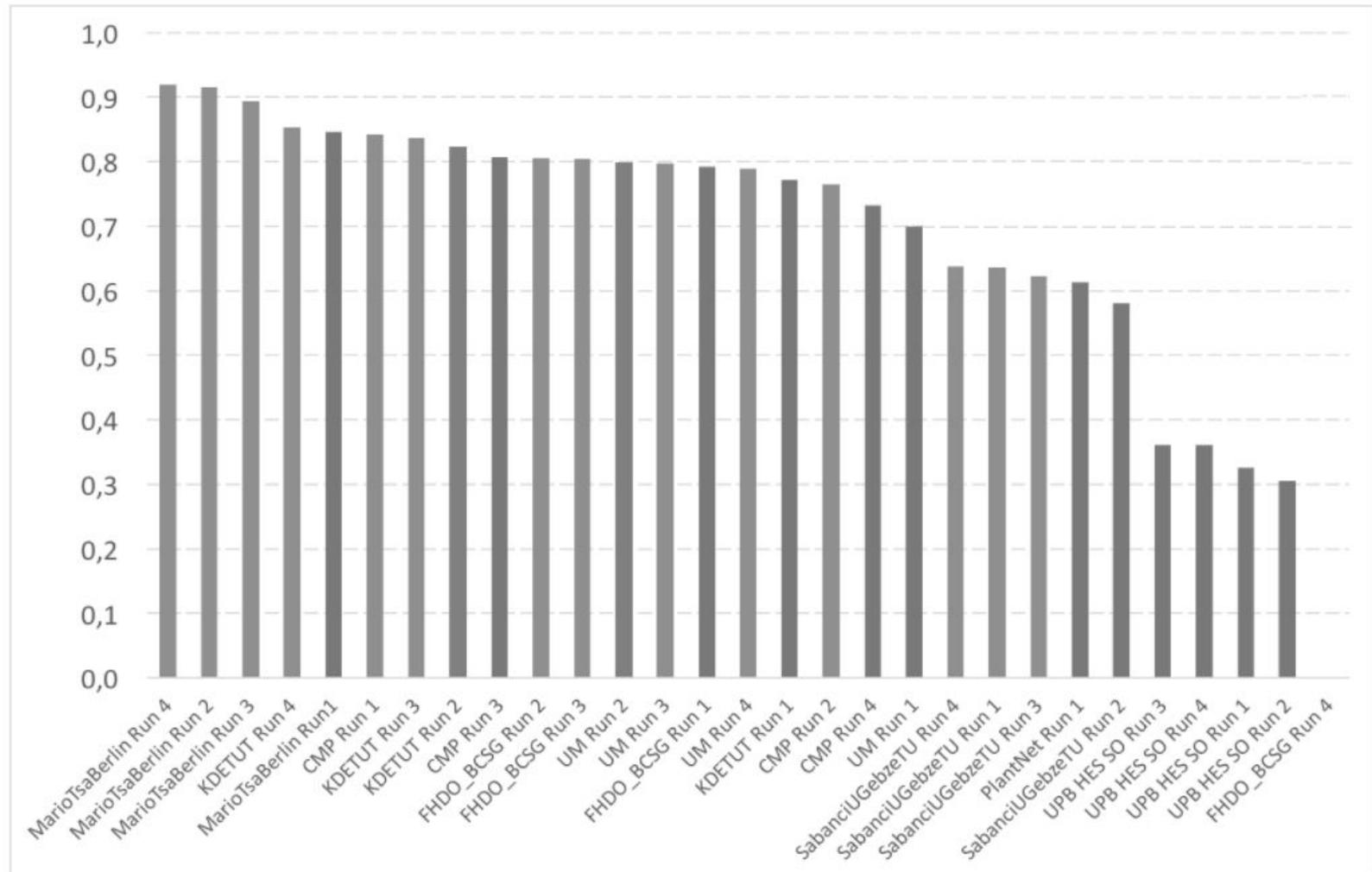
80 registered teams -> 8 teams / 29 methods



	Team	Models	Train datasets	Best MRR
	CMP	Inception-ResNet-v2	EOL	0.843
	FHDO BCSG	Inception-ResNet-v2	EOL+filtered(WEB)	0.806
	KDE TUT	Modified ResNet50	EOL+WEB	0.853
	Mario TSA Berlin	GoogleNet ResNet50 ResNext	EOL+WEB+PlantCLEF2016	0.92
	PlantNet	Inception v1	EOL	0.703
	Sabancı Gebze	VGGNet	EOL+filtered(WEB)	0.638
	UM	.VGGNet .Modified VGGNet (organ+species layers)	WEB	0.799
	UPB HES SO	AlexNet	EOL	0.361

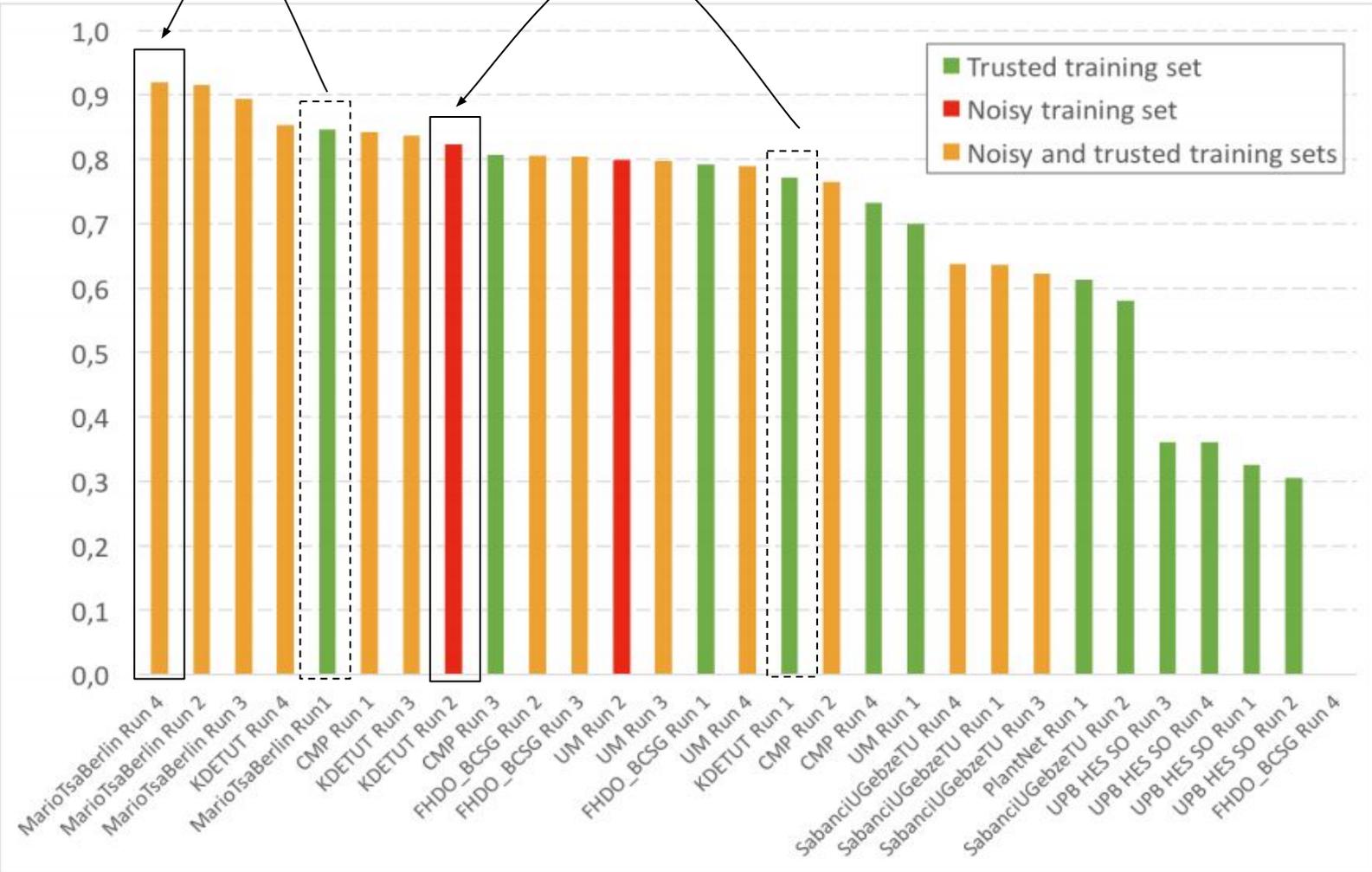
# Results

Amazing best performance on 10K species!



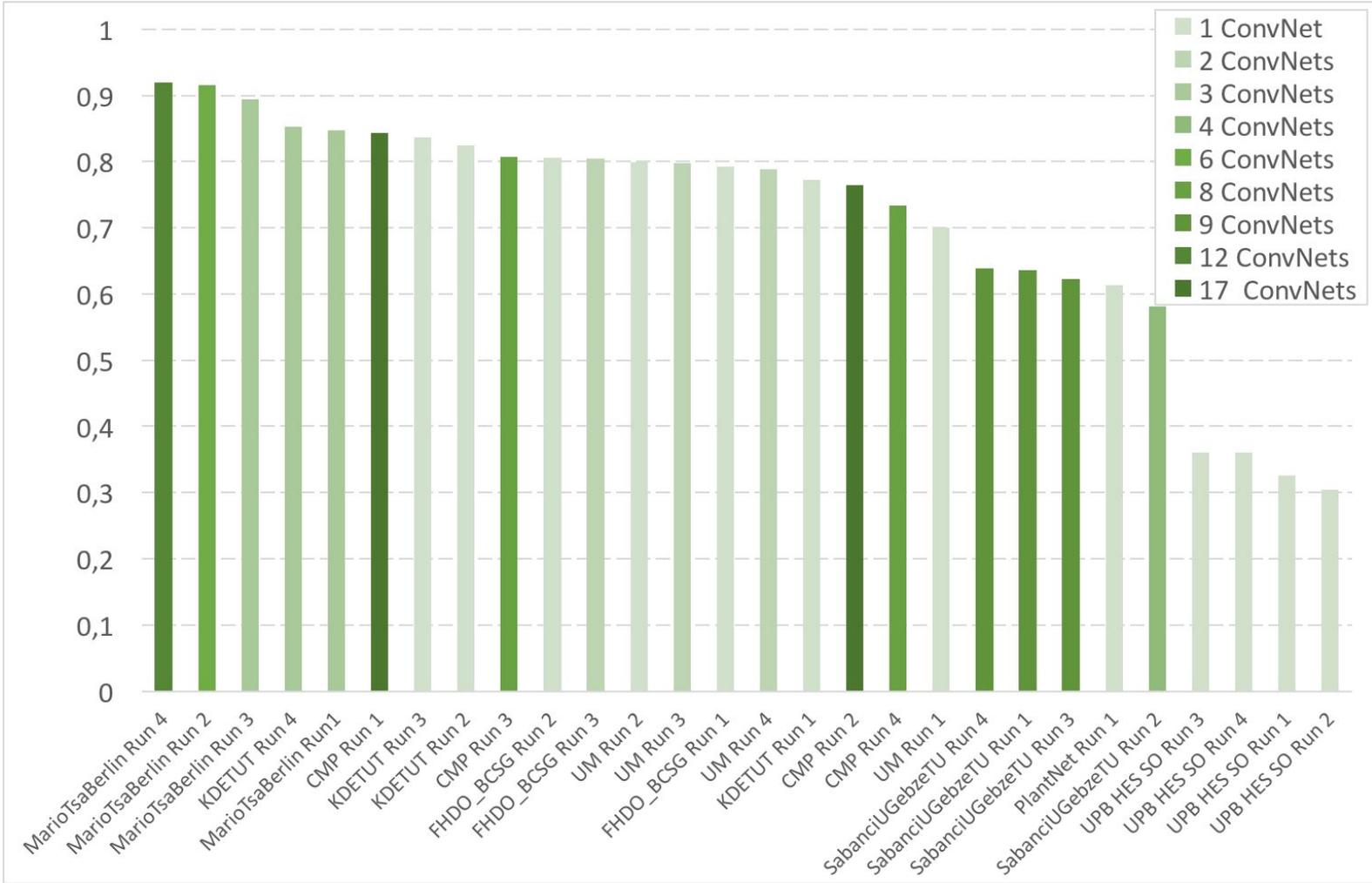
# Succeeding strategies with CNN models: **training sets**

Learning from noisy web data works better than learning from expertized data!



# Succeeding strategies with CNN models: **Bootstrap aggregating (bagging)**

the race for GPUs...

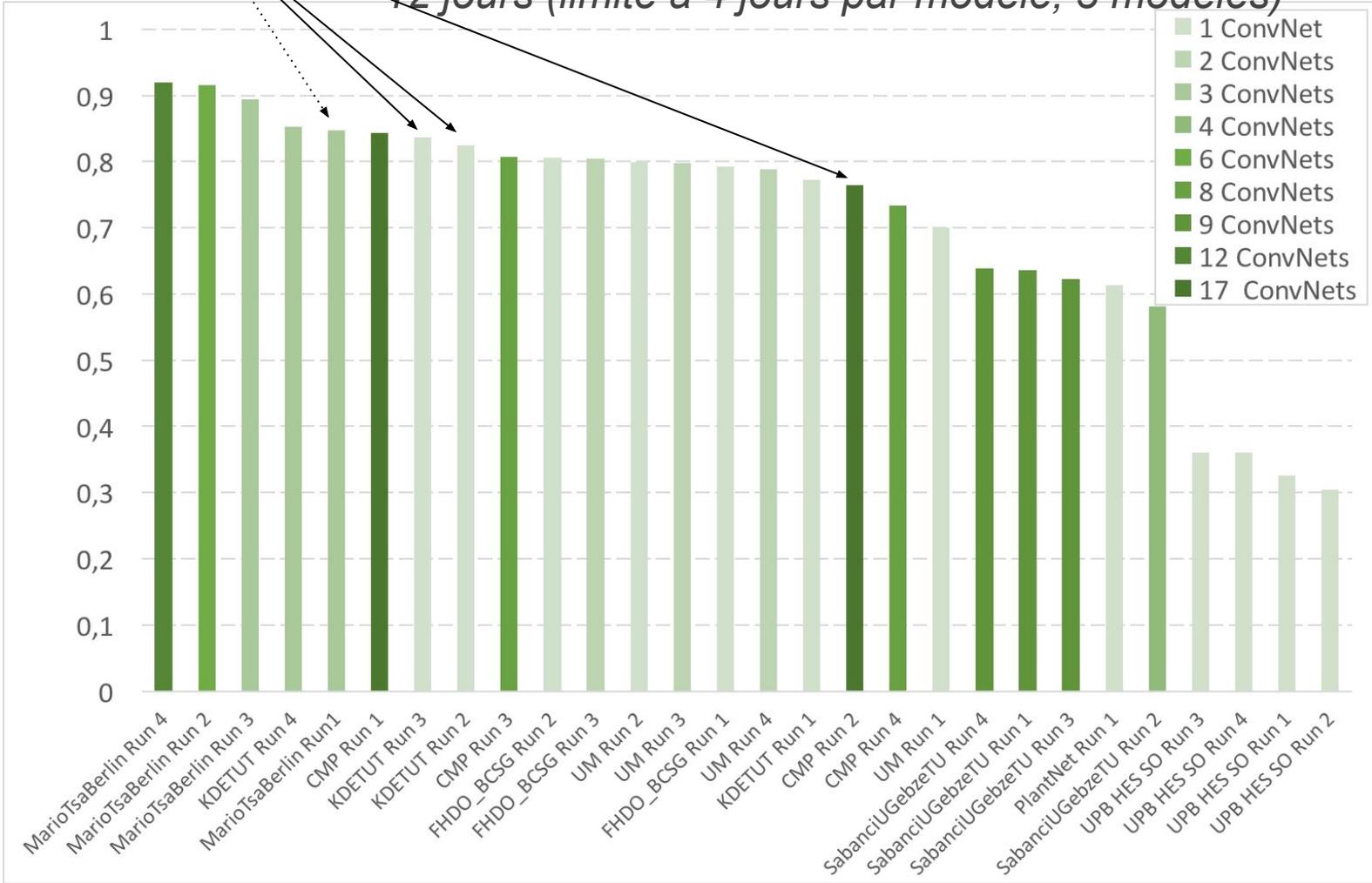


# Temps d'apprentissage



KDE TUT : ResNet50 modifié "from scratch", Tesla K40 12Go, 5 000

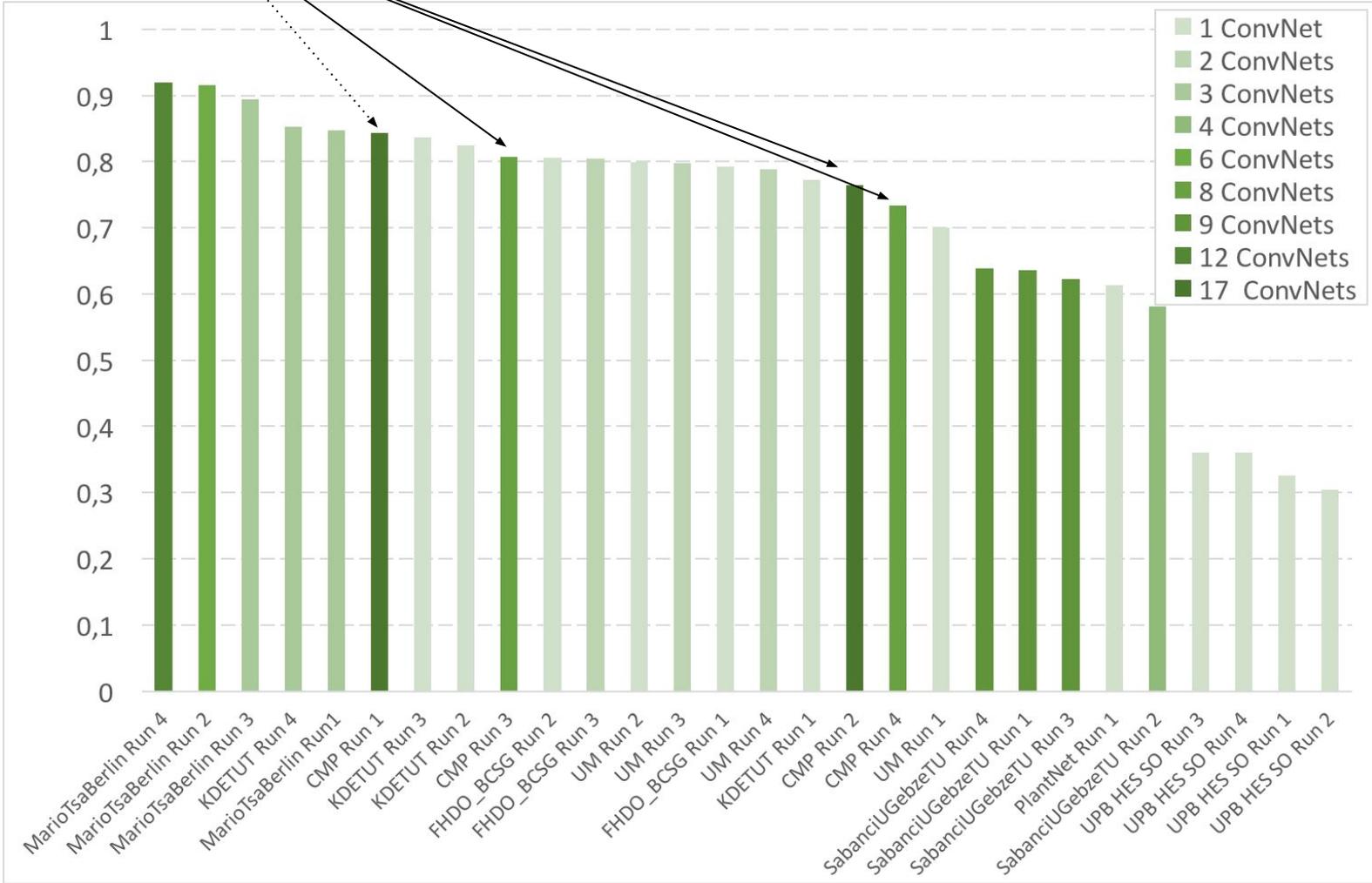
- 12 jours (limité à 4 jours par modèle, 3 modèles)



# Temps d'apprentissage

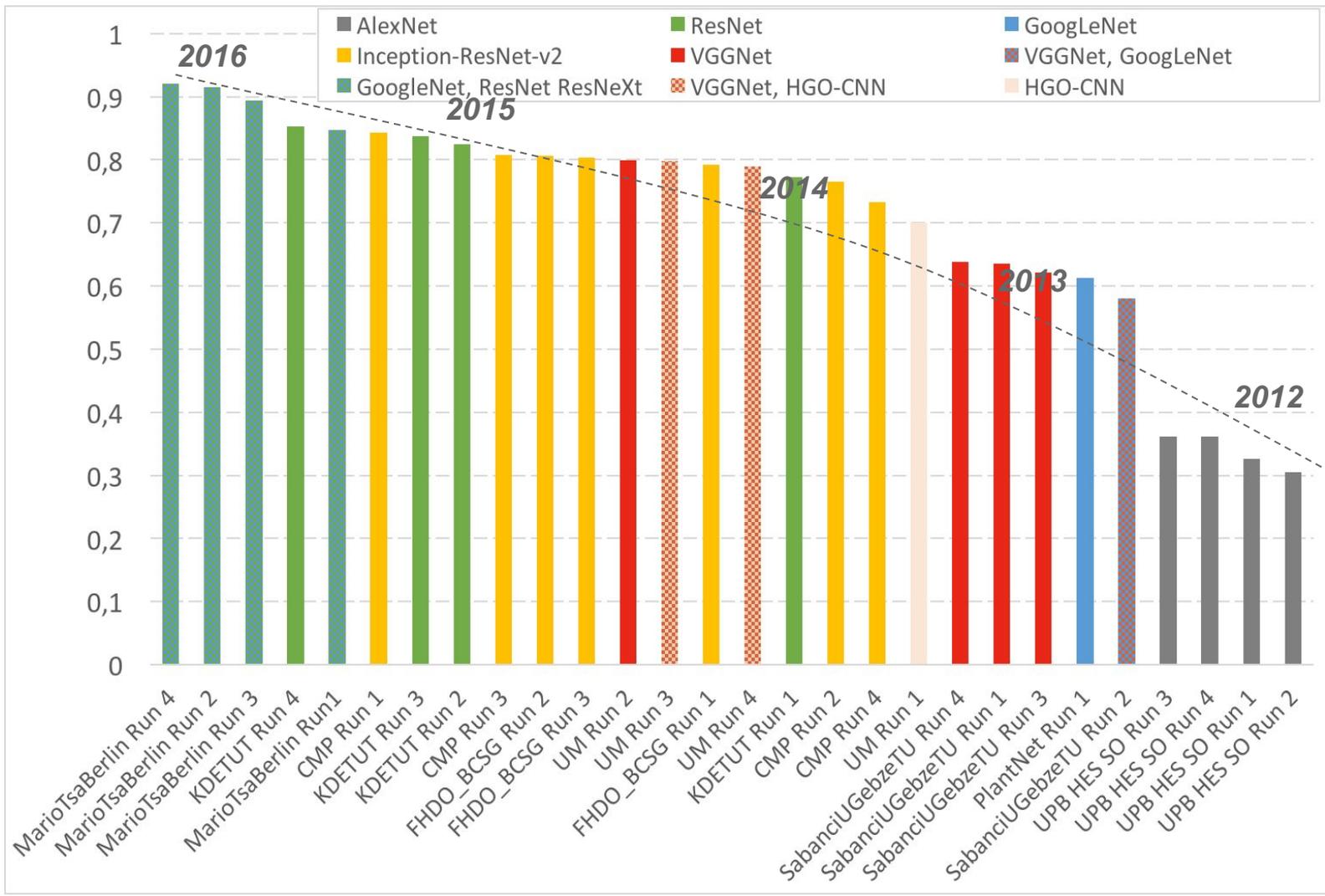


CMP : Inception-ResNet-v2 fine-tuning, Titan X 12Go, 1 300 €  
 - 60 jours (limité à 3,5 jours par modèle, 17 modèles)



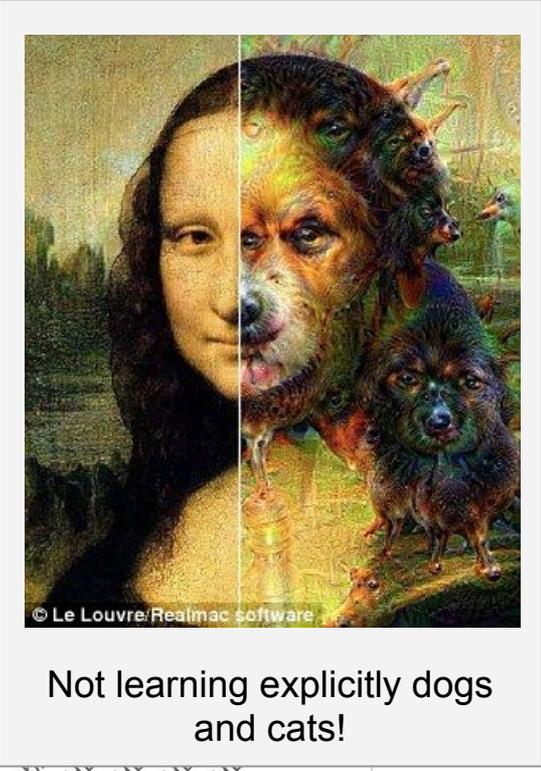
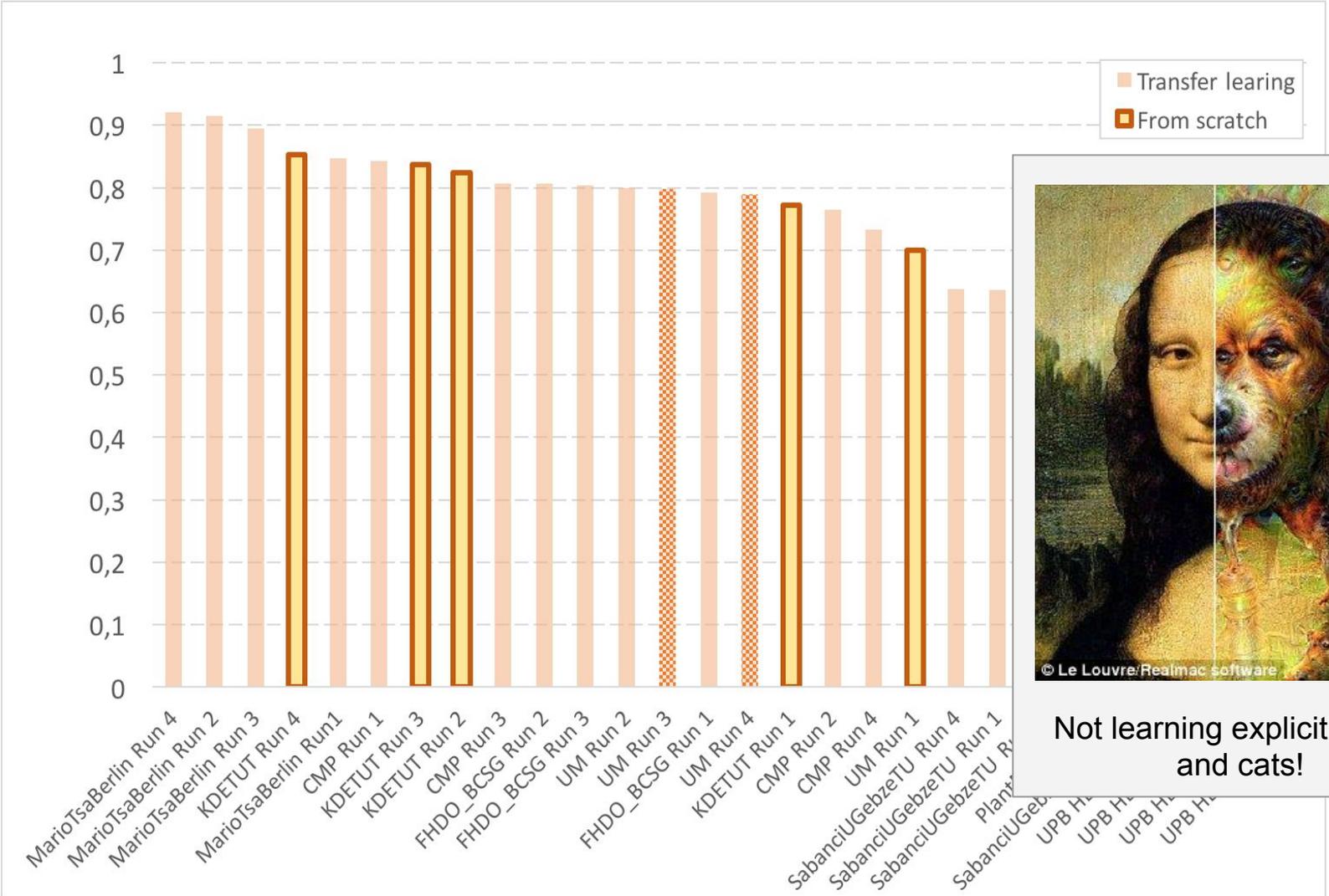
# Succeeding strategies with CNN models: architectures

... and the race for the “ultimate” model... combining new and “old” models, training strategies matter



# transfer learning vs. from scratch

10k classes with more than 1.1 millions of images for training: enough for learning from scratch?



Not learning explicitly dogs and cats!

# Succeeding strategies with CNN models: **data augmentation (on train)**

Mario TSA Berlin: more transformations (color saturation and lightness modifications)

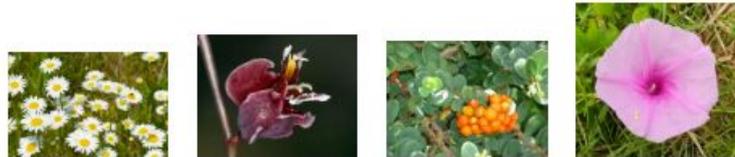


Fig. 3. Image examples with original aspect ratio



Fig. 4. Augmentation examples (MediaId 254374)



Fig. 5. Augmentation examples (MediaId 351199)



Fig. 6. Augmentation examples (MediaId 378991)

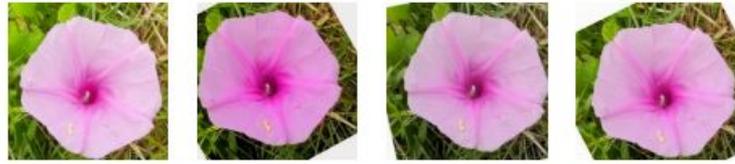
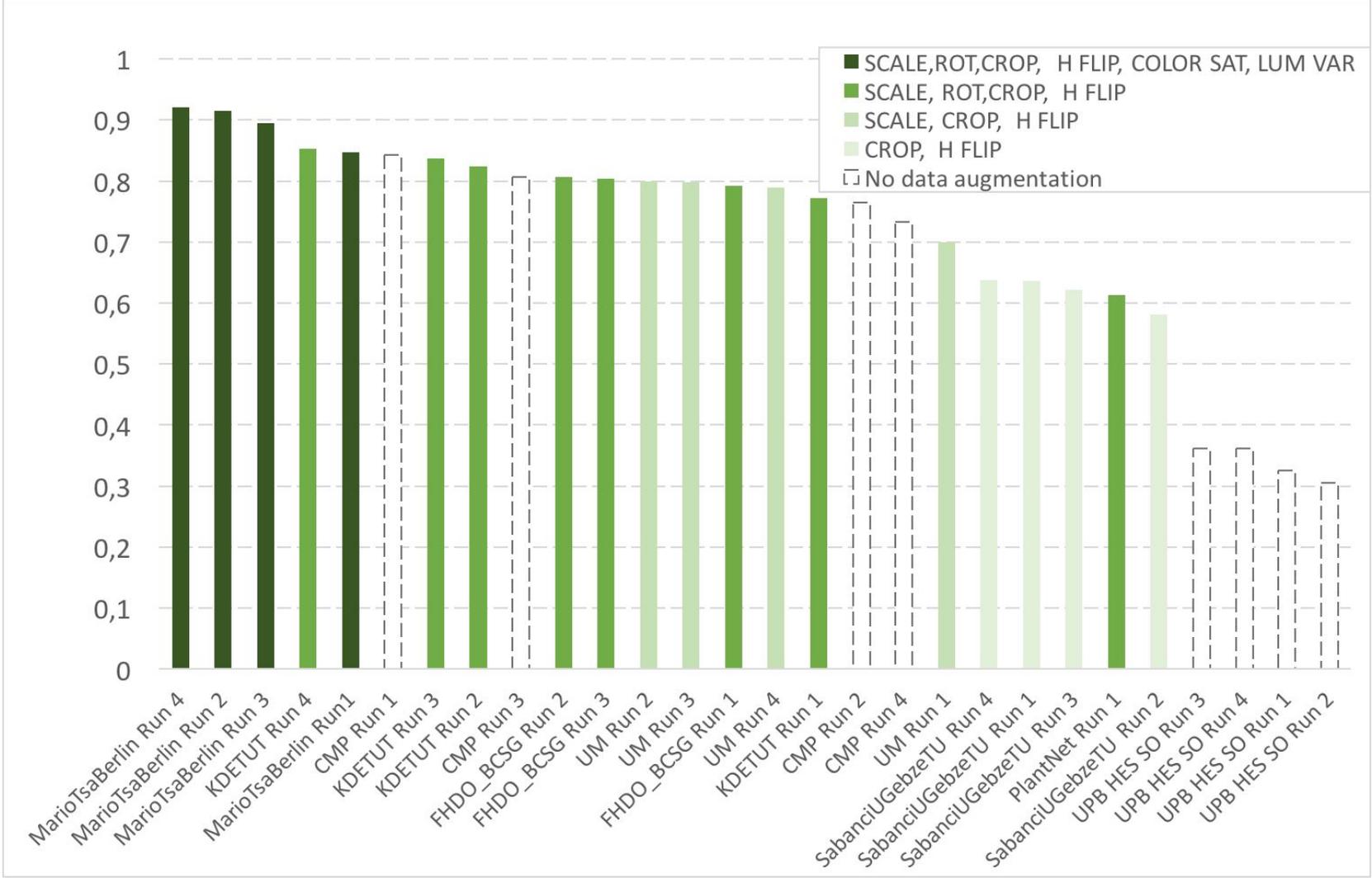


Fig. 7. Augmentation examples (MediaId 243459)

# Succeeding strategies with CNN models: **data augmentation (on train)**

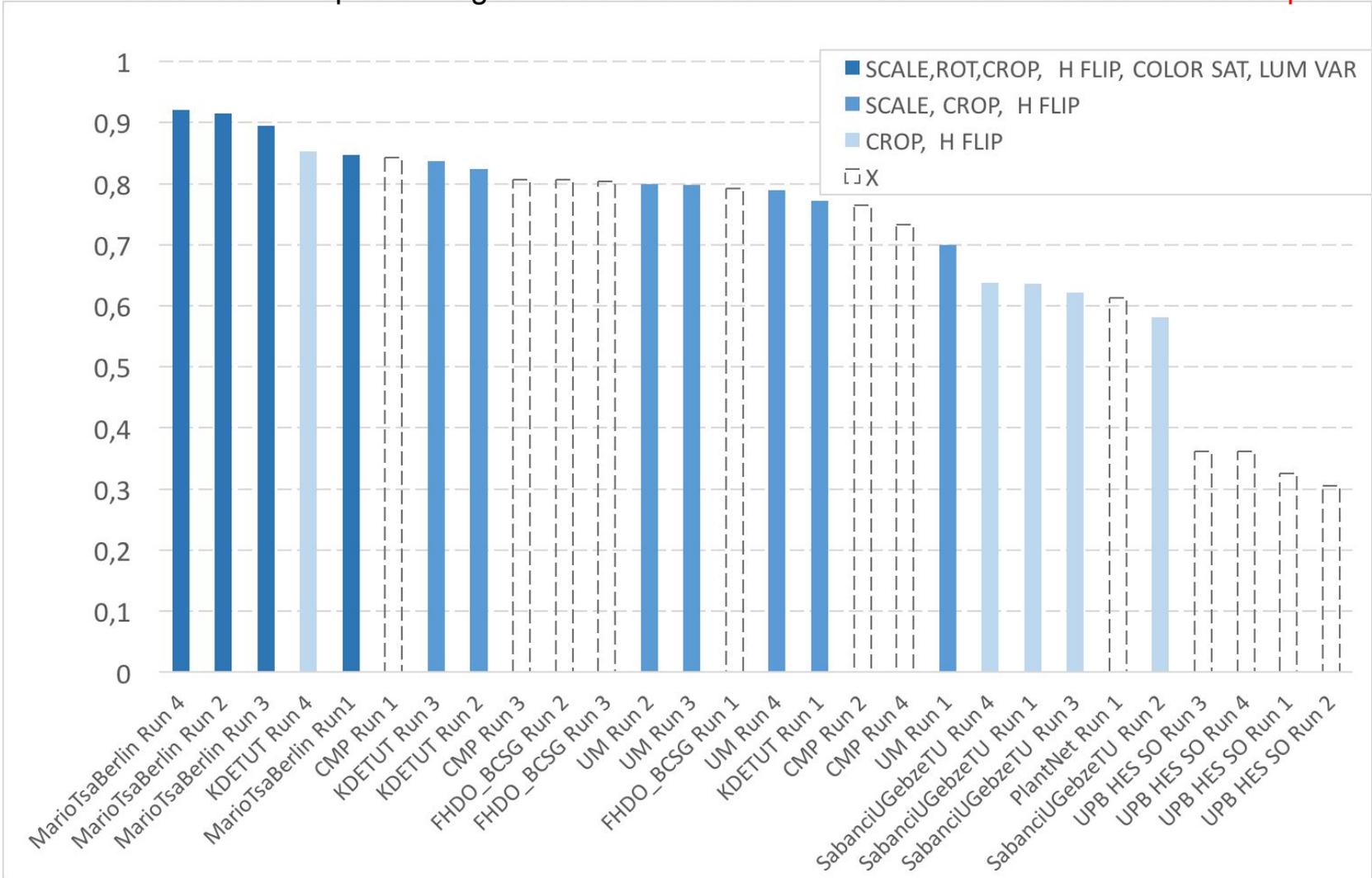
Mario TSA Berlin: more transformations (color saturation and lightness modifications)



# Succeeding strategies with CNN models: **data augmentation (on test)**

Mario TSA Berlin Run 4:

1 observation = up to 5 images x 5 transformations x 12 ConvNets = **300 combined outputs**



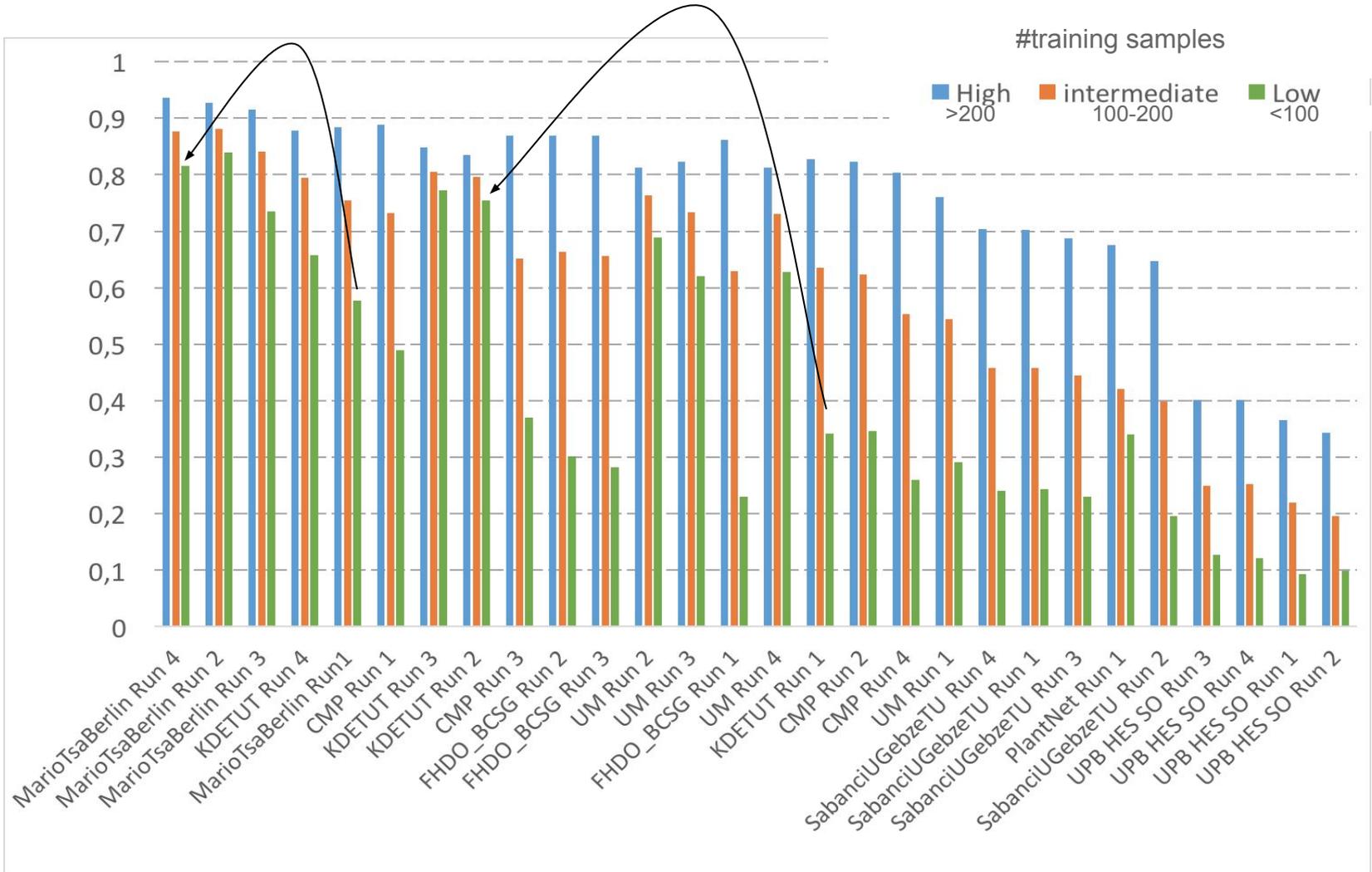
# Complementary results

the best combinations of organs?



# Results

Learning from noisy web data is more **biodiversity friendly**



# Passage à l'échelle mondiale



# Passage à l'échelle mondiale

## Catégorisation fine à très large échelle > 300 000 espèces

- Collecte et exploitation massive de données (plusieurs millions d'observations)

Sciences participatives

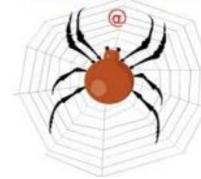


iNaturalist.org

biocollections



Crawler web



Taxonomie

The Plant List

Occurrences



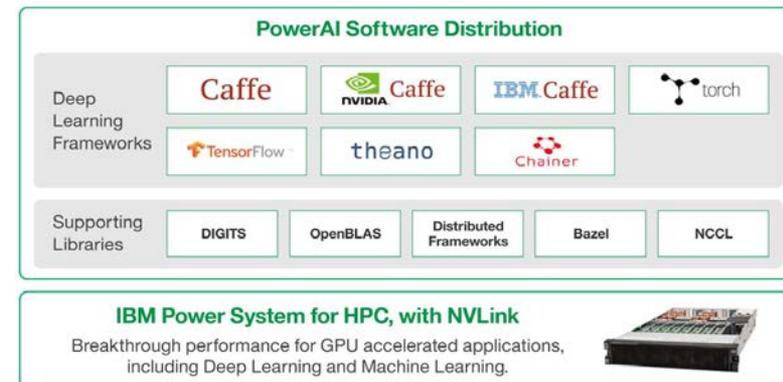
*Octobre 2017: 7,5 Millions images sur 265 000 espèces*

- Extension des capacités des ConvNets, apprentissage distribué et HPC  
*1Go le dernier FC layer pour une image, limite le parallélisme sur une seule carte, lenteur extrême*



Investissement > 1M euros ?  
48 GPUs 16 Go sur 12 serveurs IBM Minsky

IBM PowerAI Platform



# Passage à l'échelle mondiale

## Apprentissage distribué Multi-GPUs x Multi-Machines



comparaison : ImageNet 1000 classes 1,28M images "from scratch"

janvier 2016

juin 2017

août 2017

sept 2017

**15 jours**

1 GPU 4Go



**1 heure**

256 GPUs 16Go



**50 min.**

256 GPUs 16Go



**40 min.**

768 x 512 CPUs



**Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour**

Priya Goyal, Piotr Dollár, Ross Girshick, Peter Noordhuis, Lukasz Wesolowski, Adam Kyrola, Andrew Tulloch, Yangqing Jia, Kaiming He

Facebook

**Abstract**

Deep learning shines with large neural networks and large datasets. However, larger networks and larger datasets result in longer training times that impede research and development progress. Distributed synchronous SGD offers a potential solution to this problem by dividing SGD minibatches over a pool of parallel workers. To make this scheme efficient, the per-worker workload must be large, which implies exponential growth in the SGD minibatch size. In this paper, we empirically show that on the ImageNet dataset large minibatches cause optimization difficulties, but when these are addressed the trained networks exhibit good generalization. Specifically, we show no loss of accuracy when training with large minibatch sizes up to 256 images. To achieve this result, we adopt a linear scaling rule for adjusting learning rates as a function of minibatch size and develop a new warmup scheme that overcomes optimization challenges early in training. With these simple techniques, our Caffe2-based system trains ResNet-50 with a minibatch size of 256 on 256 GPUs in one hour while matching small minibatch accuracy. Using commodity hardware, our implementation achieves ~90% scaling efficiency when moving from 8 to 256 GPUs. This system enables us to train visual recognition models on internet-scale data with high efficiency.

minibatch size

validation error

Figure 1. ImageNet top-1 validation error vs. minibatch size. Error range of standard deviation is shown. We present a simple and general technique for scaling distributed synchronous SGD to minibatches of up to 50 images while maintaining the top-1 error of small minibatch training. For all minibatch sizes we set the learning rate as a linear function of the minibatch size and apply a simple warmup phase for the first few epochs of training. All other hyperparameters are kept fixed. Using this simple approach, accuracy of our models is invariant to minibatch size (up to at least 50 images). Our technique enables a linear reduction in training time with ~90% efficiency as we scale to large minibatch sizes, allowing us to train an accurate ResNet-50 model in 1 hour on 256 GPUs.

minimization [8, 10, 27]. Moreover, this pattern generalizes to non-ImageNet and non-visual classification tasks.

PowerAI DDL

Minik Cho, Ulrich Finkler, Sameer Kumar, David Kung, Vaibhav Saxena, Dheeraj Sreedhar

IBM Research

August 8, 2017

**Abstract**

As deep neural networks become more complex and input data sets grow larger, it can take days or even weeks to train a deep neural network to the desired accuracy. Therefore, distributed Deep Learning as a massive scale is a critical capability, since it offers the potential to reduce the training time from weeks to hours. In this paper, we present a software-hardware co-optimized distributed Deep Learning system that can achieve near-linear scaling up to hundreds of GPUs. The core algorithm is a multi-ring communication pattern that provides a good tradeoff between latency and bandwidth and adapts to a variety of system configurations. The communication algorithm is implemented as a library for easy use. This library has been integrated into TensorFlow, Caffe, and Torch. We train ResNet-101 on ImageNet 22K with 64 IBM Power® S822LC servers (256 GPUs) in about 7 hours to an accuracy of 33.8% validation accuracy. Microsoft's ADAM [10] and Google's DataFlow [11] results did not reach 30% validation accuracy for ImageNet 22K. Compared to Facebook's recent paper [1] on 256 GPU training, we use a different communication algorithm, and our combined software and hardware system offers better communication overhead for ResNet-50. A PowerAI DDL enabled version of Torch completed 50 epochs of training on ResNet 50 for 1K classes in 50 minutes using 64 IBM Power® S822LC servers (256 GPUs).

**1 Introduction**

Deep Learning has become the de facto technique for an increasing number of cognitive applications, including vision, speech and language translation. The success is driven by the availability of an enormous volume of data and advances in deep neural networks, but it is these enabling factors

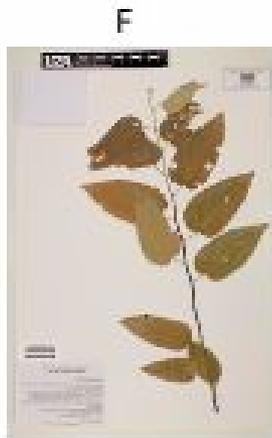
Achieving Deep Learning Training in less than 40 Minutes on ImageNet-1K & Best Accuracy and Training Time on ImageNet-22K & Places-365 with Scale-out Intel® Xeon®/Xeon Phi™ Architectures

SURF/SARA CPU, HPC, machine learning, Supercomputing, Supercomputing 04 SEP 2017

Previous post

Valeriu Codreanu, Ph.D., SURF/SARA  
Damian Podareanu, SURF/SARA  
Vikram Saleter, Ph.D., Intel Corp.

**Retour d'expérience sur la plateforme Genci/CNRS 48 GPUs (accès pendant 2 mois):**  
 tensorflow (slim) MAL documenté pour la distribution }  
 DDL-Tensorflow NON documenté, lent au démarrage } DDL-Caffe?  
 beaucoup de restrictions, fermé, pas de mise à jour de tensorflow et caffe



# Identification d'espèces dans les images d'herbiers

## Premiers travaux exploitants les ConvNets

1000 espèces (idem PlantCLEF 2015)

Top1: 0,8 !



## Evaluation de la capacité à transférer la connaissance visuelle:

- d'un herbier à l'autre, d'une flore à l'autre  
1000 sp. France -> 255 sp. Costa-Rica
- d'un type d'image à l'autre?
  - Herbier -> scan de feuille
  - Herbier -> photos terrain

	C-R	Via France
Top1	0,703	0,745

*Résultats moins concluants, les écarts visuels sont trop importants*



*2 mois  
impact élevé*



In the top 5% of all research outputs scored by Altmetric



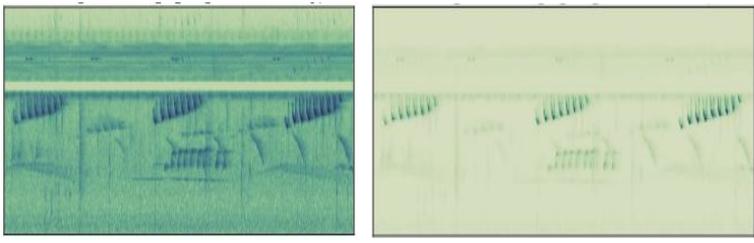
# LifeCLEF Bird Identification Task 2017

The supremacy of deep learning

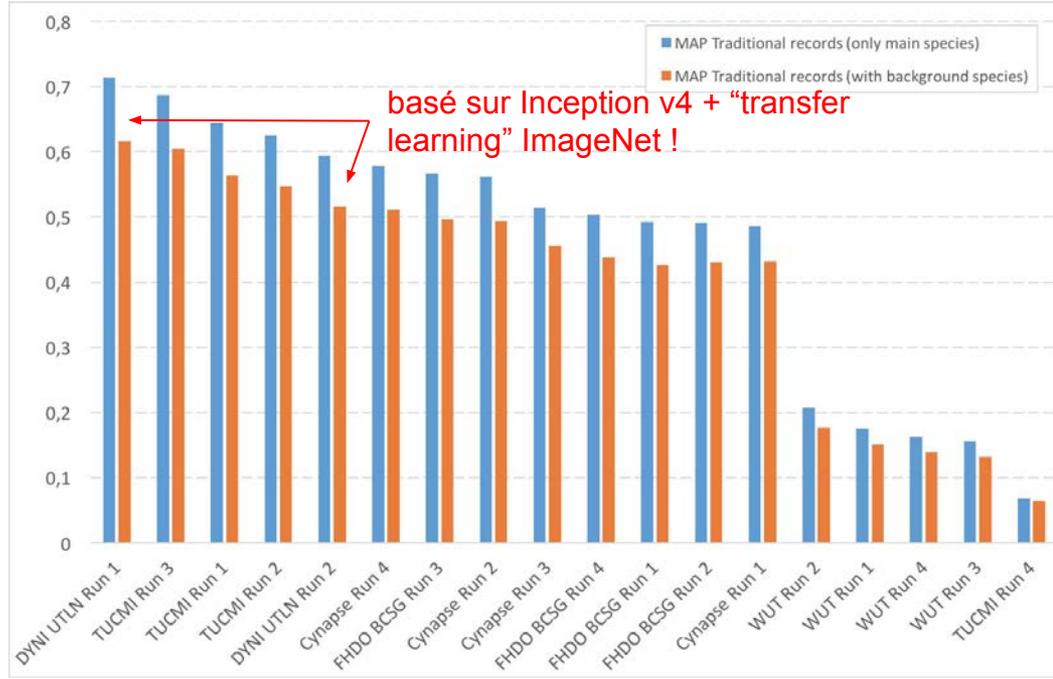


1 500 espèces centrée sur l'Amazonie  
54 000 enregistrements sonores

Signal 1D -> spectrogrammes temps fréquence



Convergence : ConvNets état de l'art peuvent apprendre à partir de spectrogrammes



### Data augmentation:

- Time shift
- Pitch shift
- Summing records of same species (multiple birds)
- Adding Noise (based one extracted noise segments)

# Conclusion

- ❖ Garder le contact avec l'état de l'art (ca va vite !)
- ❖ Il y aurait tant à explorer
  - maladie des plantes
  - identification d'adventices juveniles
  - autres règnes en interaction avec les plantes (oiseaux, insectes)
  - visualisation & construction de clé d'identifications
  - fusion de données hétérogènes
- ❖ Mais besoin de beaucoup de données d'apprentissage, de matériel, de réflexion et de temps

**Merci !**