
The Lab of CLEF dedicated to biodiversity data

# Overview of the PlantCLEF 2020 Task

## *Cross-domain plant classification (from herbariums to field photos)*

Hervé Goëau[1,2], Alexis Joly[2], Pierre Bonnet[1]

[1] Cirad, Umr AMAP, Montpellier, France

[2] Inria, Zenith team, LIRMM, Montpellier, France

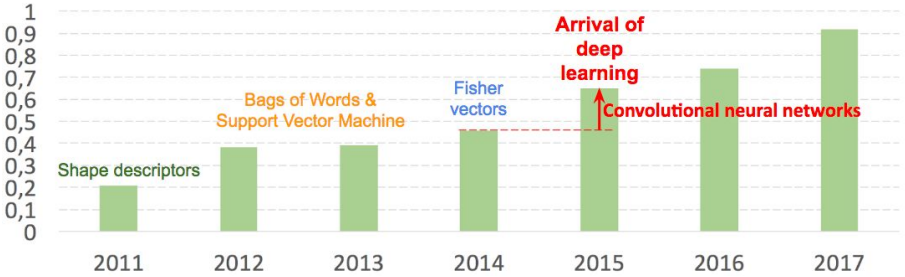
# PlantCLEF: a long term evaluation

nowadays automated systems perform well in **temperate regions**

- deep learning
- big data

*Top1 PlantCLEF 2018: 0,88*

| | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 |
|---|---|---|---|---|---|---|---|
| Species | 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 |
| Nb. of particip. | 8 | 11 | 12 | 22 | 15 | 16 | 17 |
| **Best perf.** | **0,209** | **0,38** | **0,393** | **0,456** | **0,652** | **0,742** | **0,92 !** |




PlantNet Plant Identification
plantnet-project.org Education
PEGI 3

PictureThis: Identify Plant, Flower, Weed and More
Glority LLC Education
PEGI 3

PlantSnap - FREE plant identifier app
PlantSnap, Inc. Education
PEGI 3

Flora Incognita - automated plant identification
Technische Universität Ilmenau Education
PEGI 3

Seek by iNaturalist
iNaturalist Education Education
PEGI 3 Family Friendly

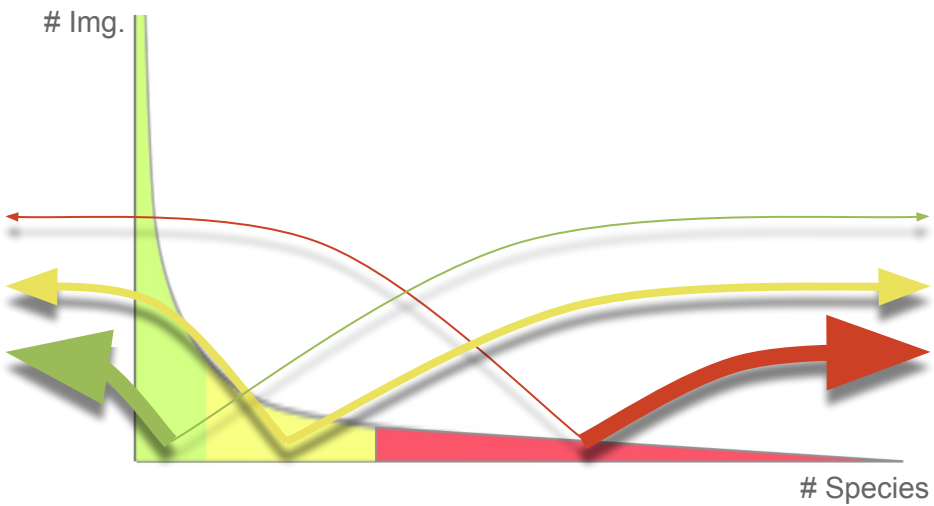
# PlantCLEF: a long term evaluation

nowadays automated systems perform well in **temperate regions**

- deep learning
- big data

**...but poorly in tropical regions:**

*Top1 PlantCLEF 2018: 0.88*




*Plant biodiversity long tail distribution*

*Top1 PlantCLEF 2019: 0.25*

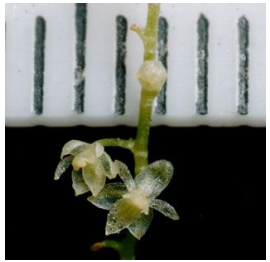
# Data deficient tropical countries

remote isolated areas

plant in canopy

many species in a genus

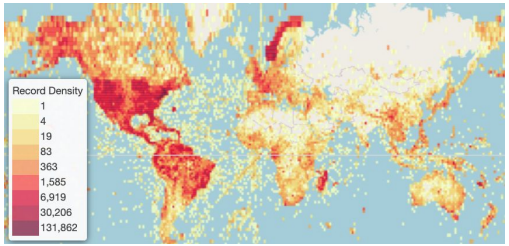
very small sp.

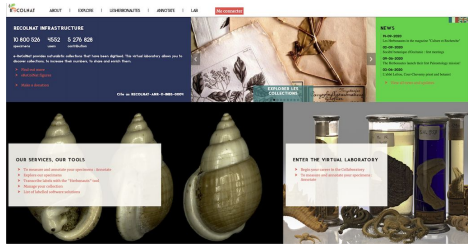


**But potentially millions of underexploited digitized herbarium sheets collected over centuries**

iDigBio
Integrated Digitized Biocollections

RECOLNAT



# A cross domain classification task

One **training** sample in domain of **herbariums**


One **test** sample in domain of **field photos**

**An “easy” example**

*- brownish, dry and matt content VS shiny green leaves*

*- but the "obovate" shapes and the nervation should match both domains*

*Unonopsis stipitata Diels*

# A cross domain classification task

One **training** sample in domain of **herbariums**


One **test** sample in domain of **field photos**

**A more difficult ex.**

*- global view vs close-up*

*- dry fruits vs immature fruits*

*Inga acrocephala Steud.*

# A cross domain classification task

One **training** sample in domain of **herbariums**


One **test** sample in domain of **field photos**

*Strychnos cayennensis Krukoff & Barneby*

**A difficult example**

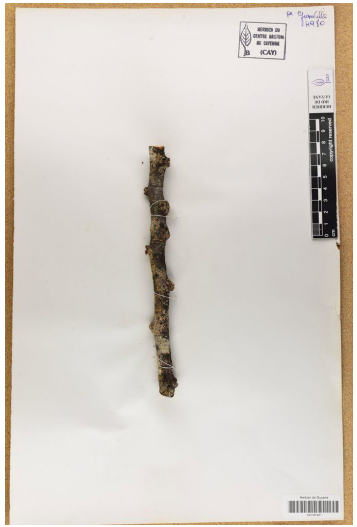
*- two branches with leaves and fruit(s)*

*- but flattened vs 3D perspective (side view of the fruit), light*

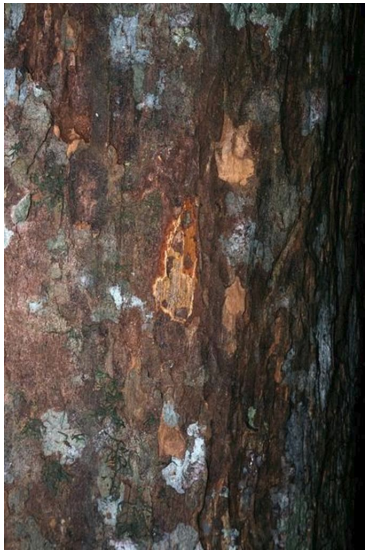
*- strips of scotch tape, several textual annotations, paperclip, envelopes ...*

# A cross domain classification task

**Training** samples in domain of **herbariums**



One **test** sample in domain of **field photos**



*Bocoa prouacensis Aubl.*

**A impossible example**

*- barks are rarely collected in herbariums*

*- plant observations with several photos of various organs may compensate the lack of information*

# A cross domain classification task

- **State of the art?**
  - **“classical” CNNs?**(ability to represent features in a common space?)
  - **vs domain adaptation approaches?**

- **Performances?**
  - **Overall?**
  - **vs genericity on rare species?** (on species with only herbarium sheets as training examples)
  - **On common species with many training photos** (do the herbarium sheets disturb here the performances?)

Vitex orinocensis Kunth

# Dataset

**TRAINING SET**

L'HERBIER IRD DE GUYANE

iDigBio Integrated Digitized Biocollections

997 species (French Guiana)

330,752 herbarium sheets

4,482 field photos (375 sp)

354 specimens on both domains

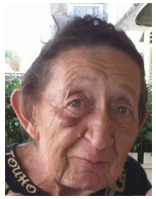
external data allowed

Life Cle · eol Encyclopedia of Life · bing · GBIF Global Biodiversity Information Facility · Google images

**TEST SET**

Two world-class experts of the Guyana flora

*Tribute to Marie-Françoise "Fanchon" Prévost*

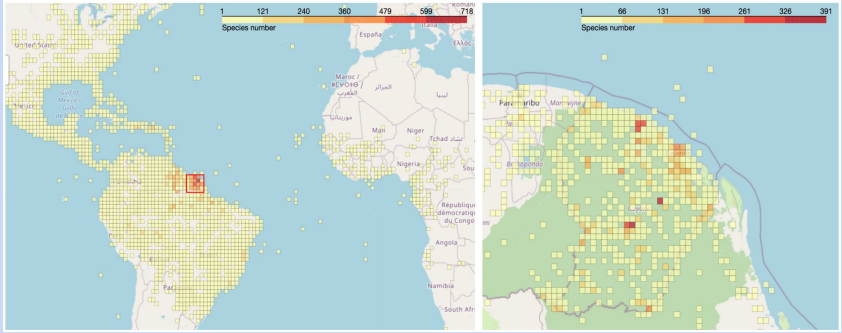


*Jean-François Molino*

IRD Institut de Recherche pour le Développement FRANCE

638 plant observations related to 3,186 field photos and 408 species

Grid density map of the dataset

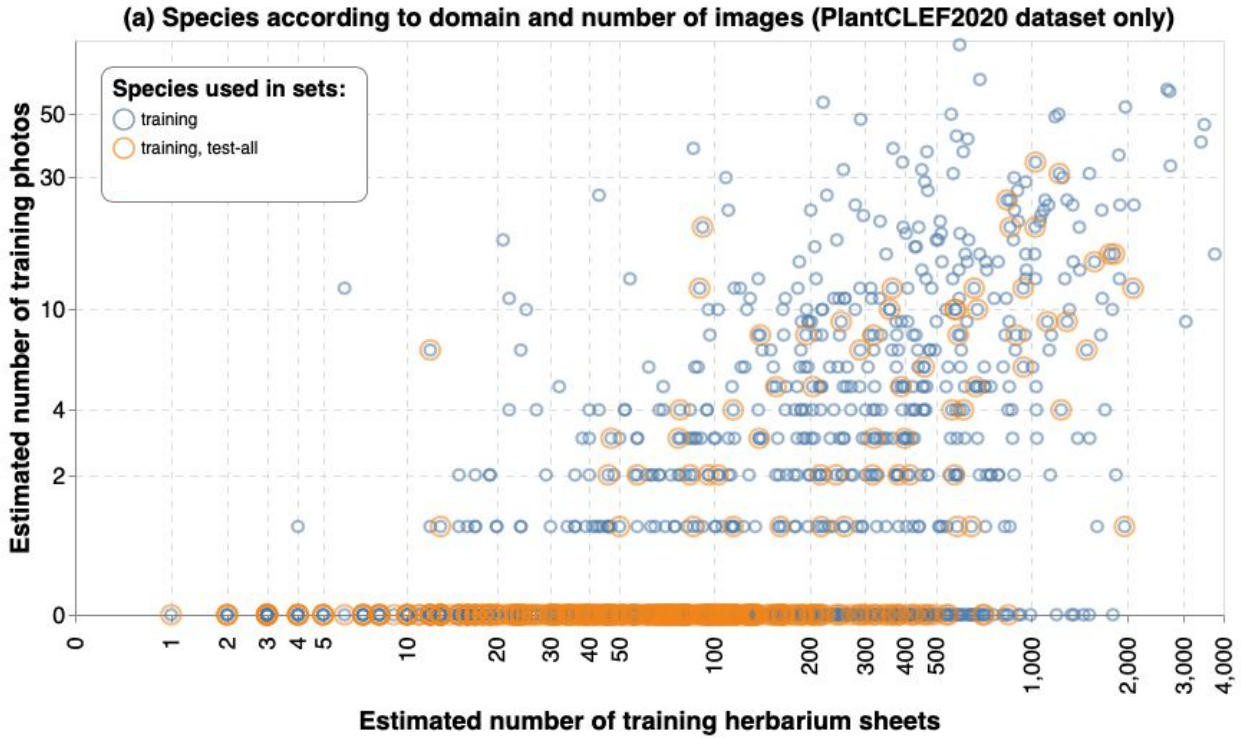


# Metric (primary)

Mean Reciprocal Rank

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

Q = 638 plant observations

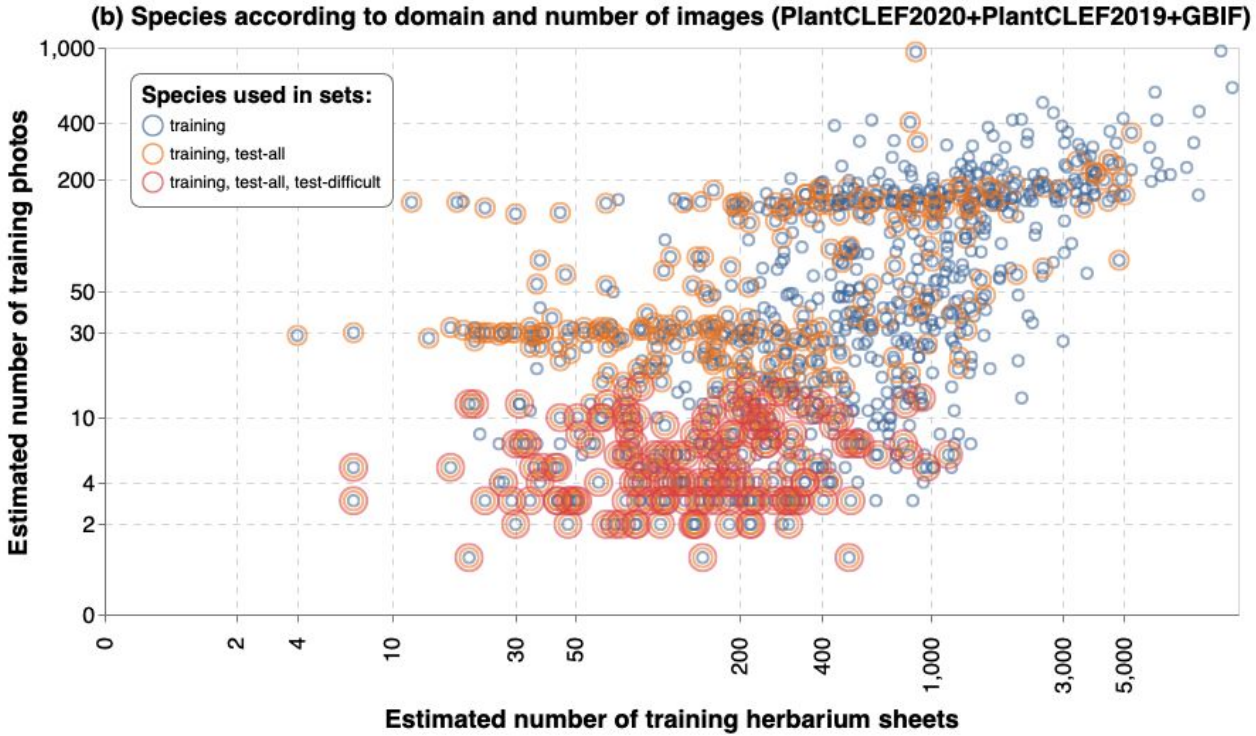


# Metric (secondary)

Mean Reciprocal Rank

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

Q = **200** plant observations **related to the most difficult species "in the world"**




Picek, L., Šulc, M., Matas, J.: Recognition of the amazonian flora by inception networks with test-time class prior estimation. In: CLEF (Working Notes) (2019)

# Participation

AIcrowd

71 registered teams
7 teams crossed the finish line
49 tested methods






SSN

UNIVERSITY OF WEST BOHEMIA

LEHIGH UNIVERSITY

NEUON

TEC | Tecnológico de Costa Rica

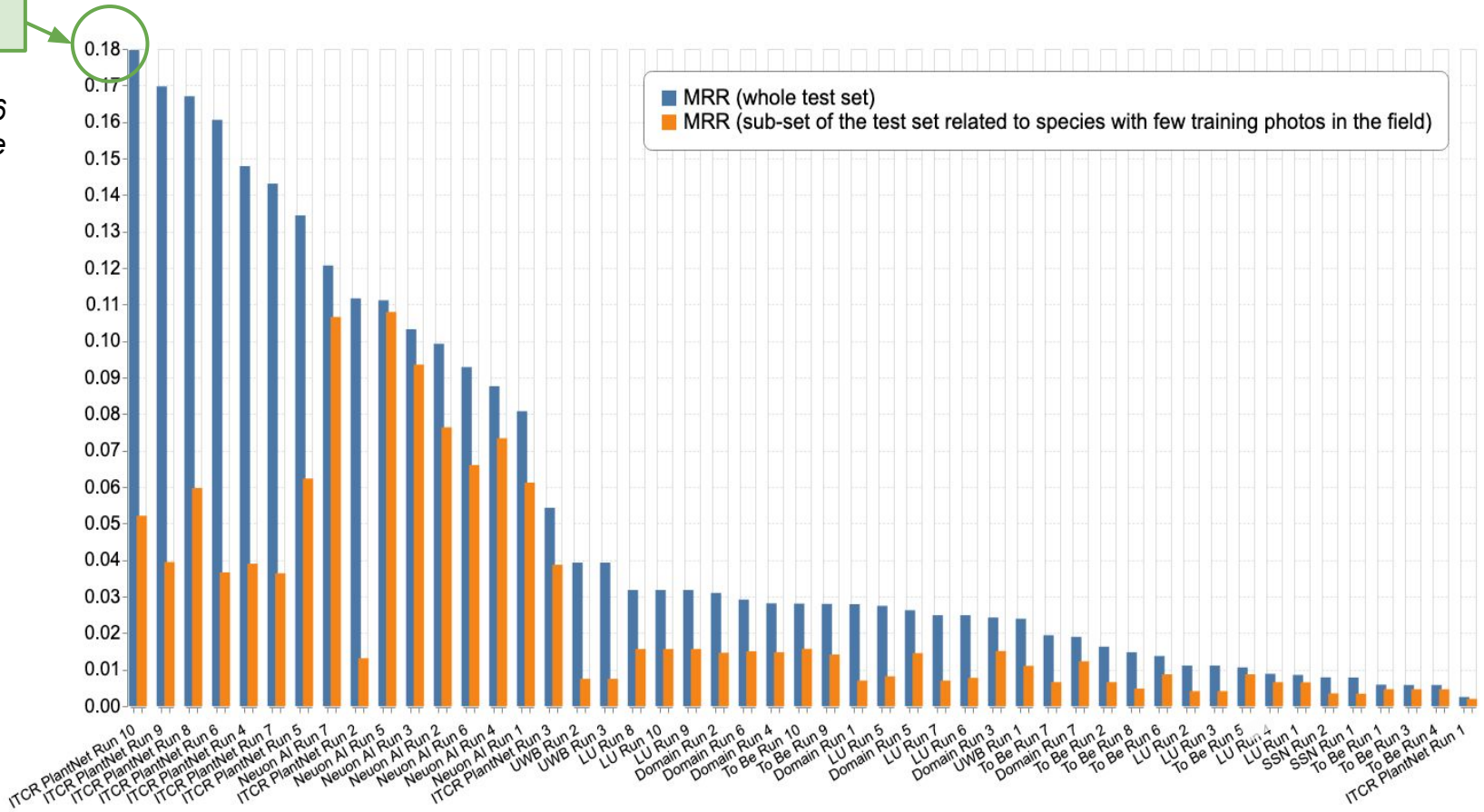
Pl@ntNet

| | | |
|---|---|---|
| Finetuned CNNs | | Domain Adaptation |
| ImageNet -> Herbarium | ResNet50 | Common agnostic feature space<br>Encoders & discriminators |
| ImageNet -> Herbarium + Photos | Inception-resnet-v2 | |
| ImageNet -> Herbarium -> Photos | inception-v4 | Mapping<br>Triplet loss & siamese network, cycada |
| With or without external data | | Auxiliary tasks (self supervision, genus & family classifiers) |
| Separate data augmentation techniques adapted to each domain | | |

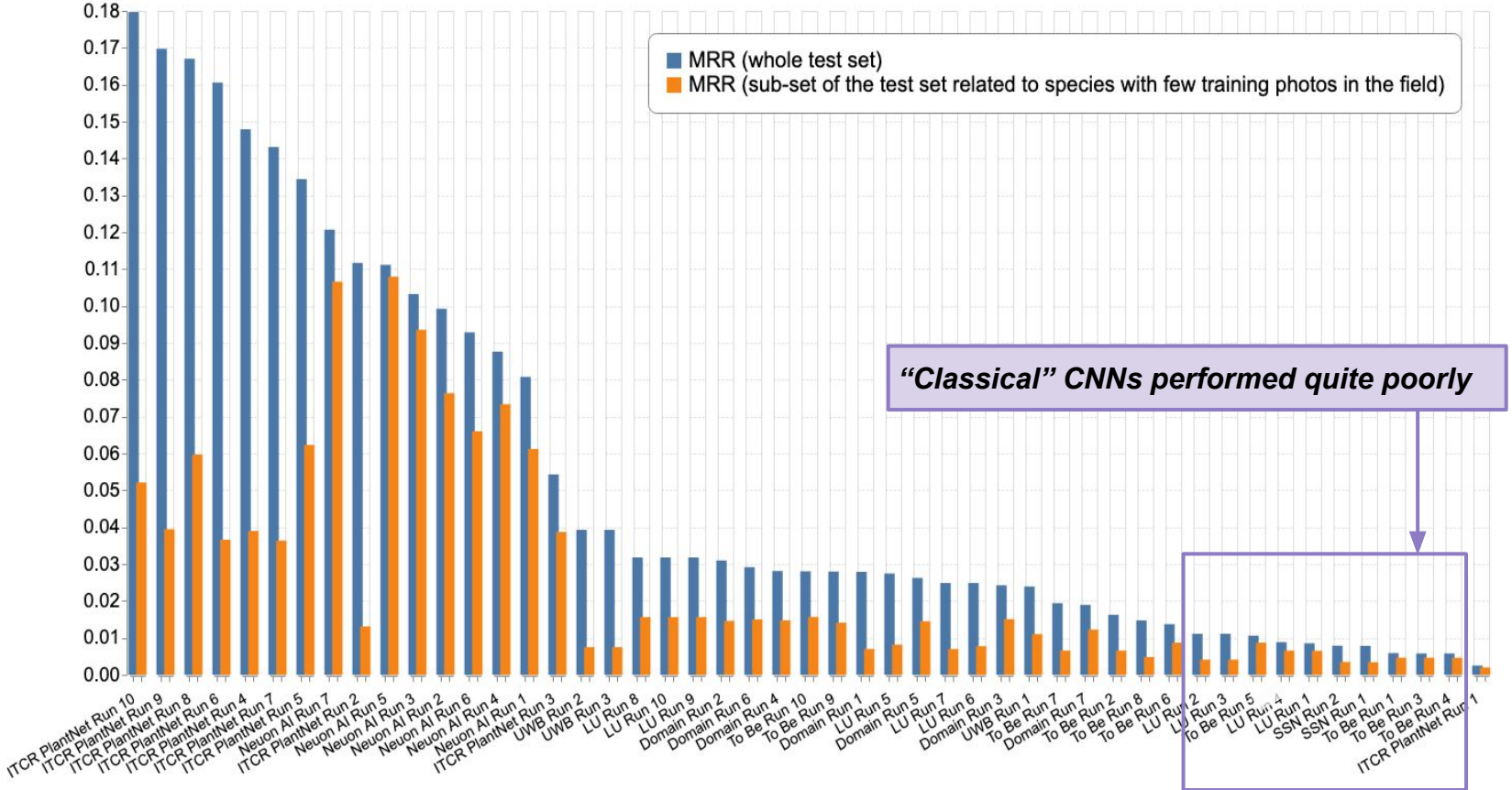
# Results:

***A difficult task***

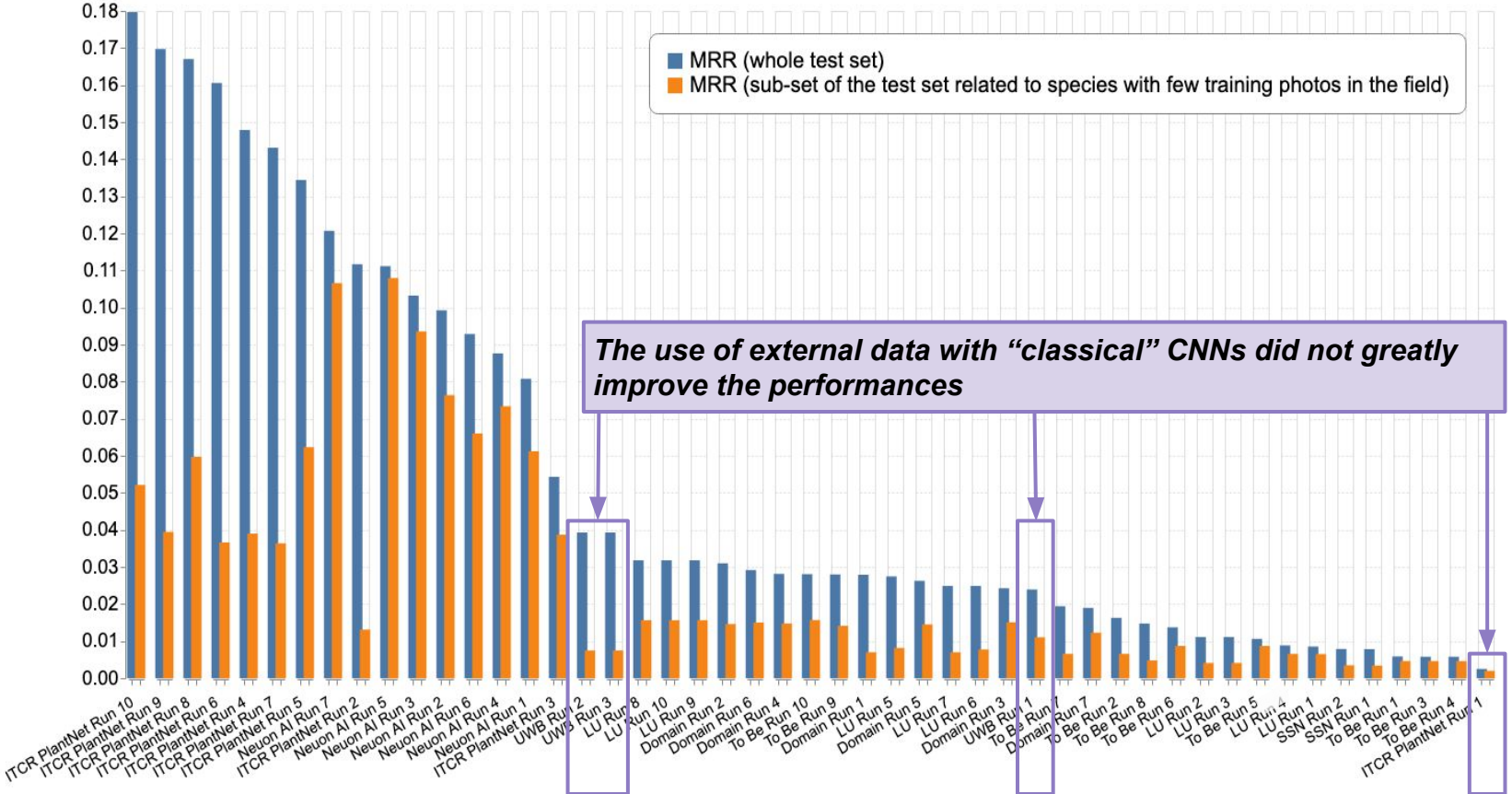
*MRR 2019: 0.376 (same flora, more species, mainly photos in the training set)*



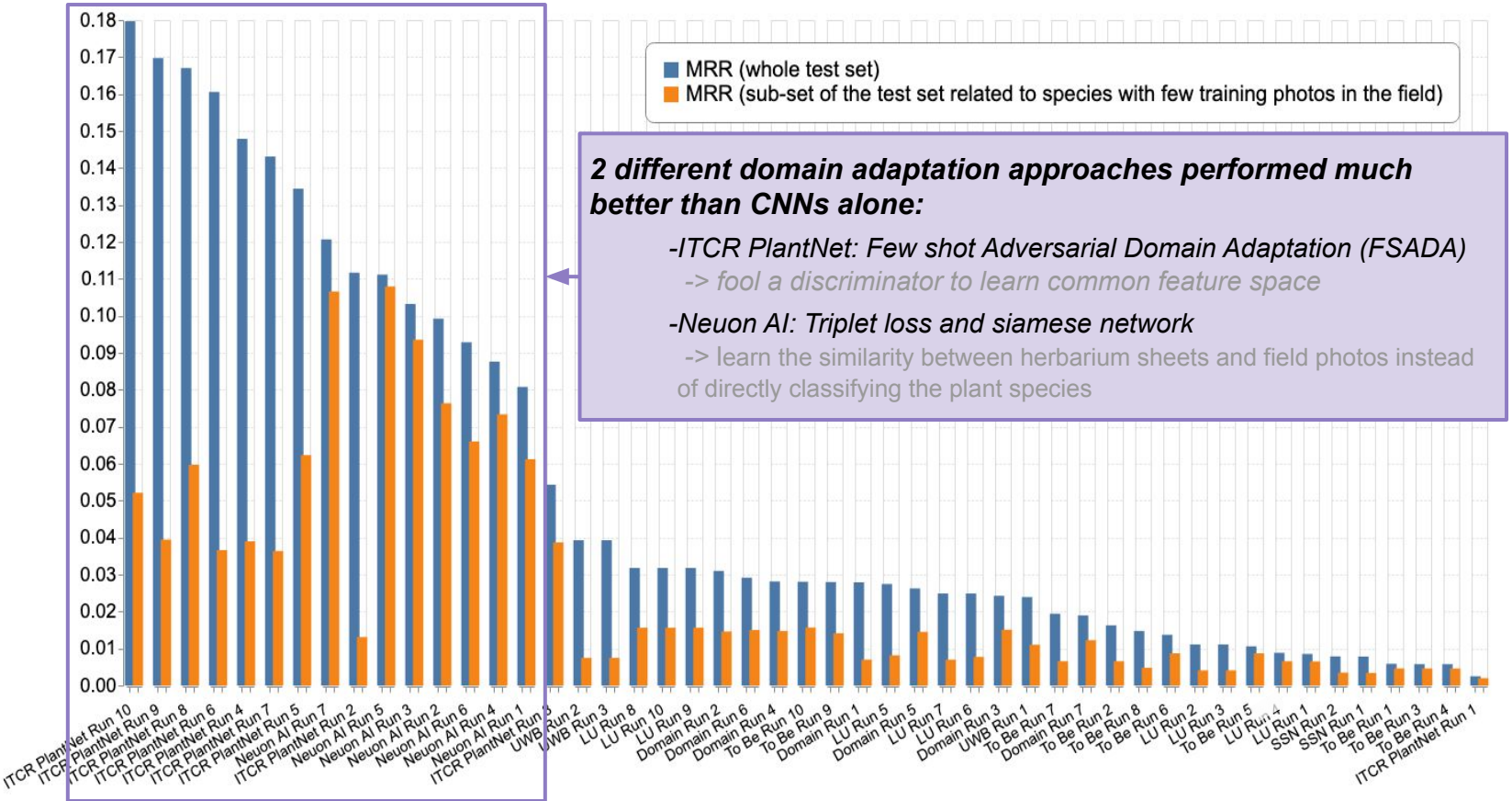
# Results: “classical” CNNs



# Results: “classical” CNNs + external data



# Results: “classical” CNNs vs domain adaptation




***2 different domain adaptation approaches performed much better than CNNs alone:***

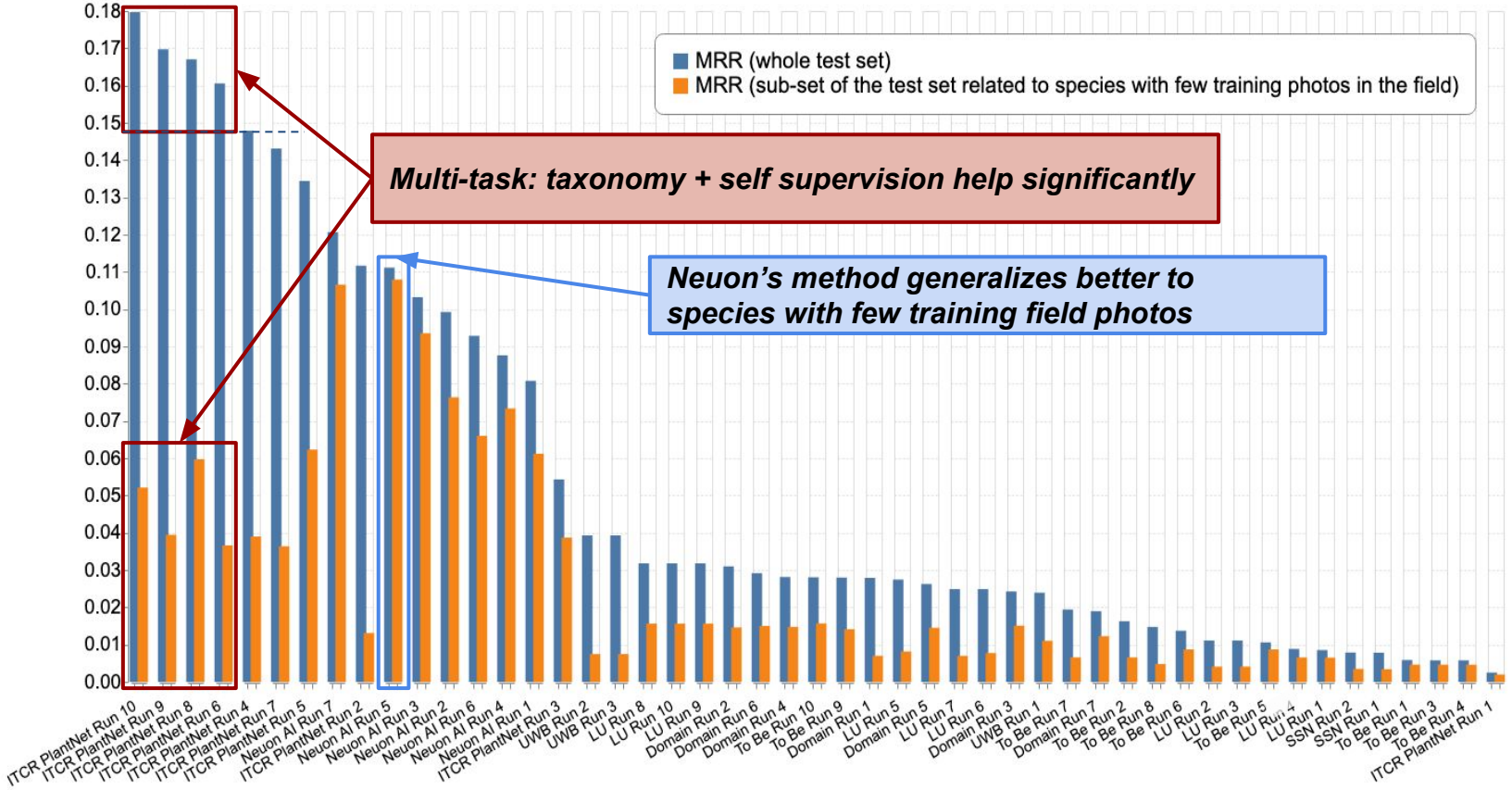
*-ITCR PlantNet: Few shot Adversarial Domain Adaptation (FSADA)*

*-> fool a discriminator to learn common feature space*

*-Neuon AI: Triplet loss and siamese network*

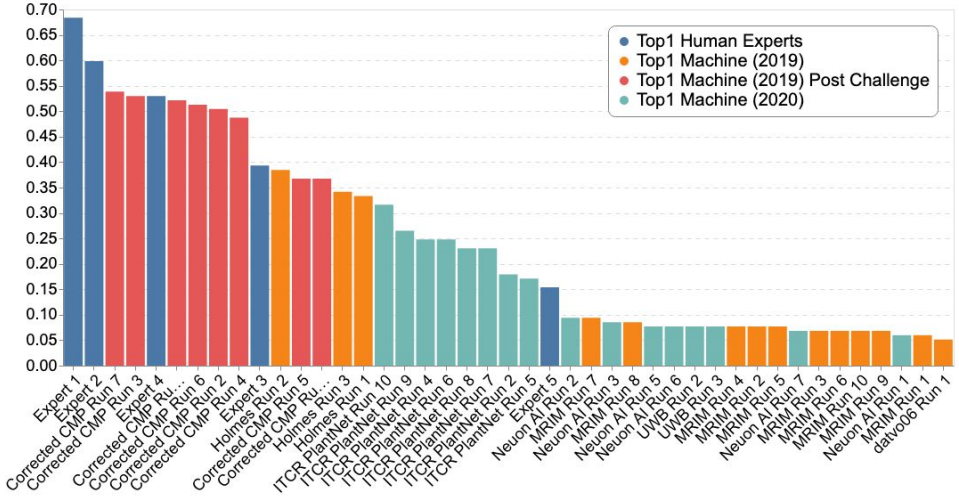
-> learn the similarity between herbarium sheets and field photos instead of directly classifying the plant species

# Results: overall performances vs genericity

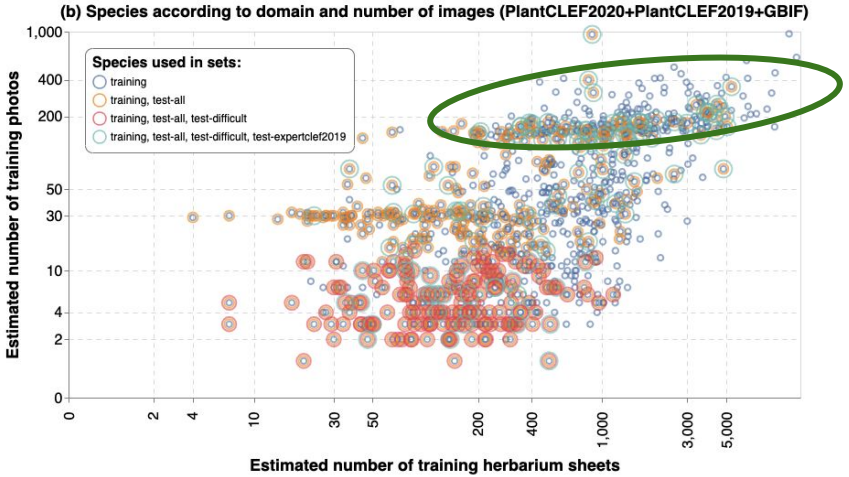


# ExpertCLEF 2019: replay

The massive use of herbarium sheets had no impact regarding the last year's results...




...since the plant observations in this test set are related to species with many training field photos




*Predictions filtered to 1k classes for fair comparison*

# Conclusions

The most **difficult task** in all PlantCLEF editions

**Domain adaptation** approaches outperformed “classical” CNNs

**Rare species**: have to deal with a compromise between **genericity** and overall performance

**Common species**: herbarium sheets not really profitable for the species with many training photos

-> still searching for an efficient universal technique!

# Thank you

agropolis fondation