

# Extracting Trait Data from Digitized Herbarium Specimens using Deep Convolutional Networks

Sohaib Younis, Marco Schmidt, Claus Weiland, Stefan Dressler, Susanne Tautenhahn, Jens Kattge, Bernhard Seeger, Thomas Hickler, Georg Zizka and Robert Hoehndorf

# Outline

- Motivation
- Related Work
- Overview
- Work Flow
- Image-Trait Dataset
- Deep Learning Approach
- Result
- Outlook

# Motivation

- Rapidly improving deep learning methods.
- Availability of biodiversity Big Data by large scale digitization efforts.
- Increasing demand for automated feature detection in agro-technical, conservational, ecological and climate change domains.

## Related Work

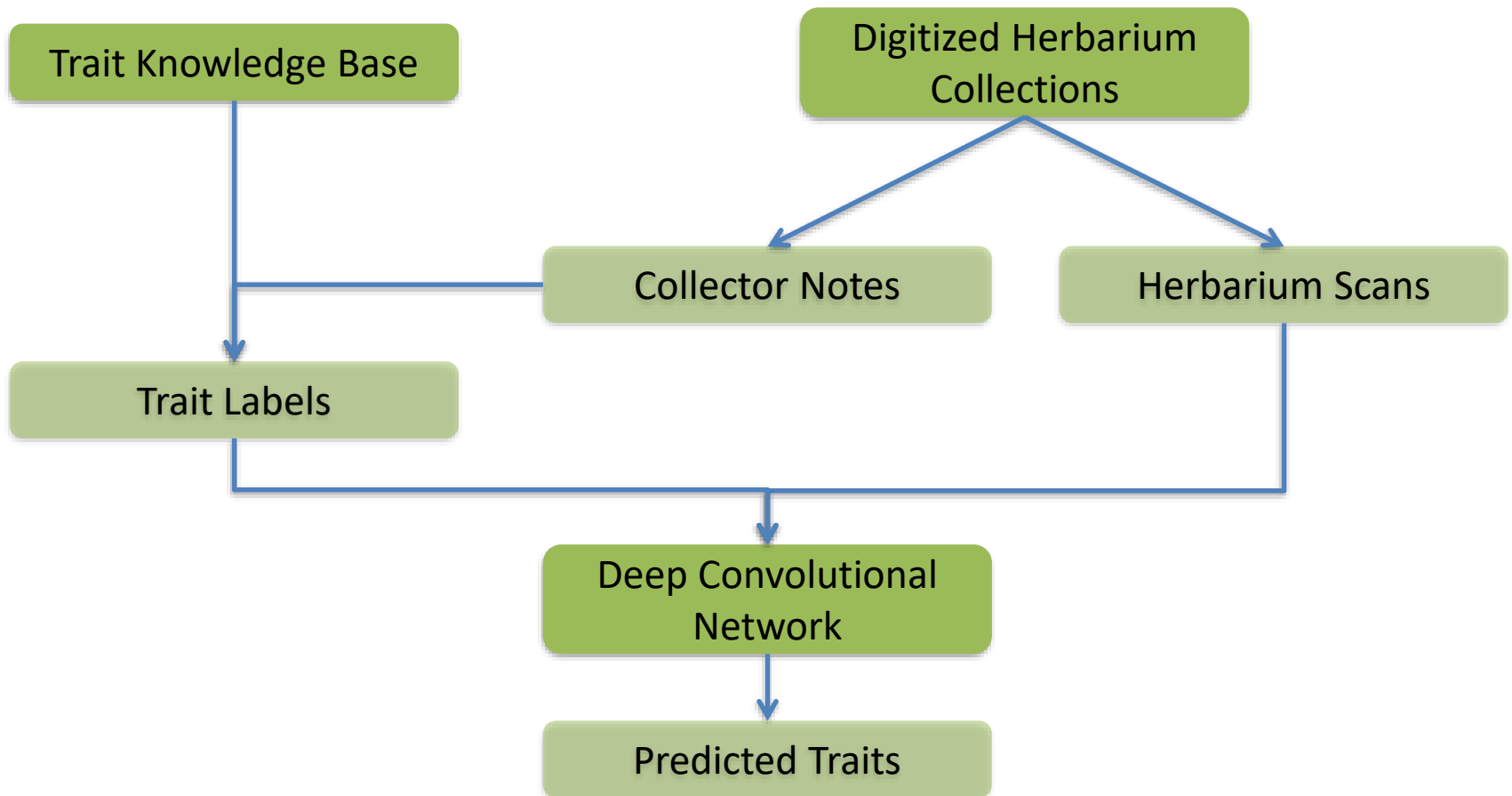
- Deep Plant Phenomics: A Deep Learning Platform for Complex Plant Phenotyping Tasks (Ubbens and Staveness 2017)
- Deep machine learning provides state-of-the-art performance in image-based plant phenotyping (Pound et al. 2017)
- Taxon and trait recognition from digitized herbarium specimens using deep convolutional neural networks (Younis et al. 2018)

## Overview

“Mobilization of trait data from digital image files by deep learning approaches.”

- Recognize traits of plants from digital herbarium scans.
- Link image dataset (GBIF) with trait knowledge base (FLOPO) to create labels for herbarium scans.
- Develop a deep learning framework for detection and prediction of traits.

# Workflow

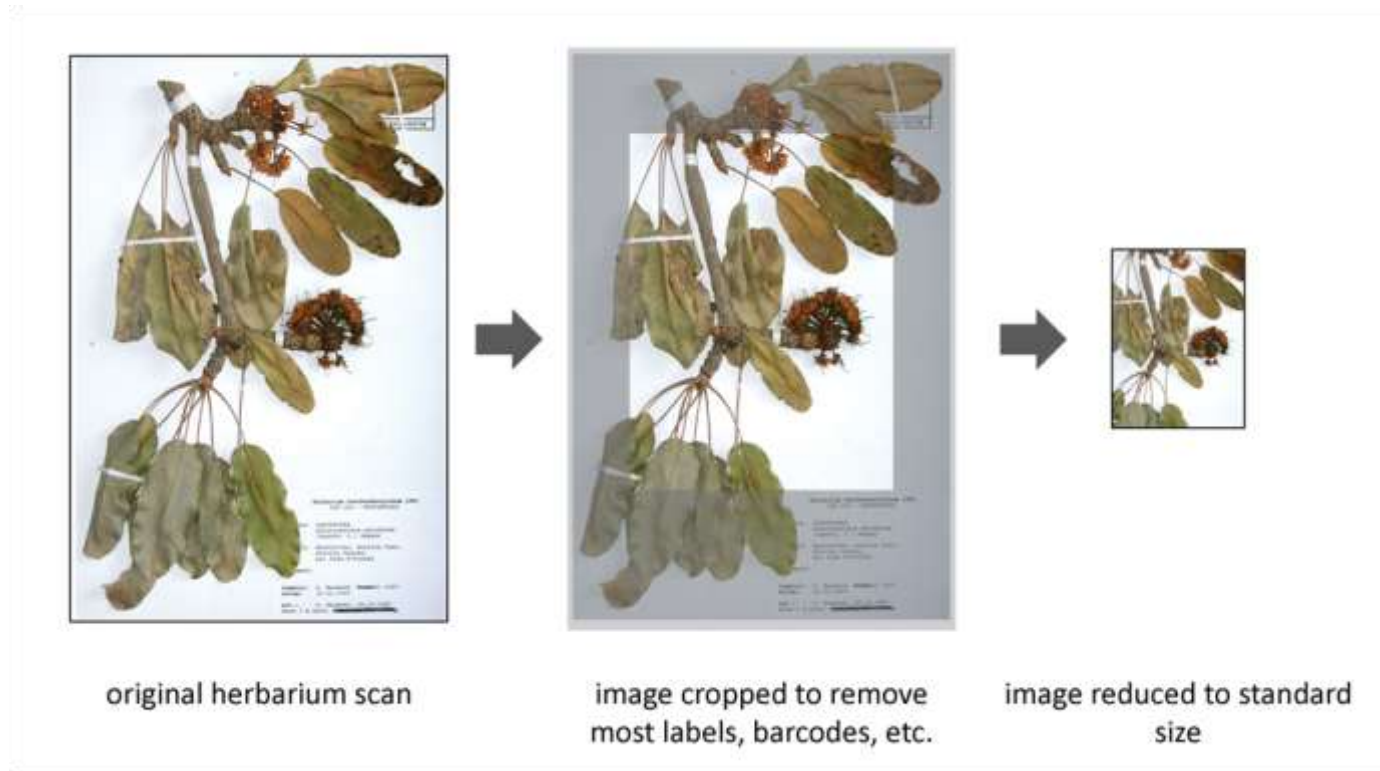


# Herbarium Collections

(via GBIF)

- Royal Botanic Gardens, Kew (K)
- Royal Botanic Garden Edinburgh Herbarium (E)
- Herbarium of Universite de Montpellier, Institut de Botanique (MPU)
- The vascular plant collection at the Herbarium of the Museum national d'Histoire Naturelle (P)

# Herbarium Scans



Herbarium scans as downloaded have been cropped and reduced to a standard size (1200 x 800) for the convolutional neural network (Younis et al. 2018)



## Extracting Plant Organs

Text mining based keyword search for fruit and flower organs / parts in collector notes.

Sample keywords:

- fruit, berry, pod, legume, capsule, nut, etc.
- flower, corolla, corona, petal, sepal, etc.

# Collector Notes and Descriptions

Text based notes taken by collectors/observers.

**Description:** Slender climber with yellow flowers

**Notes :** East facing hillside along ridge and down to roadside shelter. Deciduous tree 6 m high; white sap. Fruits green/red.

**Notes :** Evergreen forest, in shade, roadside. Tree 8 m high. Fruits green, immature.

**Description:** Evergreen shrub to 4 m tall; flowers creamy white

## Royal Botanic Garden Edinburgh Flora of Chile

Calceolariaceae

*Calceolaria glandulosa* Poepp. ex Benth.

**Locality:** Región V [Valparaíso], Prov. Petorca, Quebrada del Tigre. 32°33'32.53"S, 71°26'43.97"W. 209m

**Habitat:** Open grassy, west-facing slopes and surrounded by sclerophyllous vegetation dominated by large trees of *Beilschmedia miersii*, *Schinus latifolia*, *Aextoxicon punctatum*, *Cryptocarya alba*

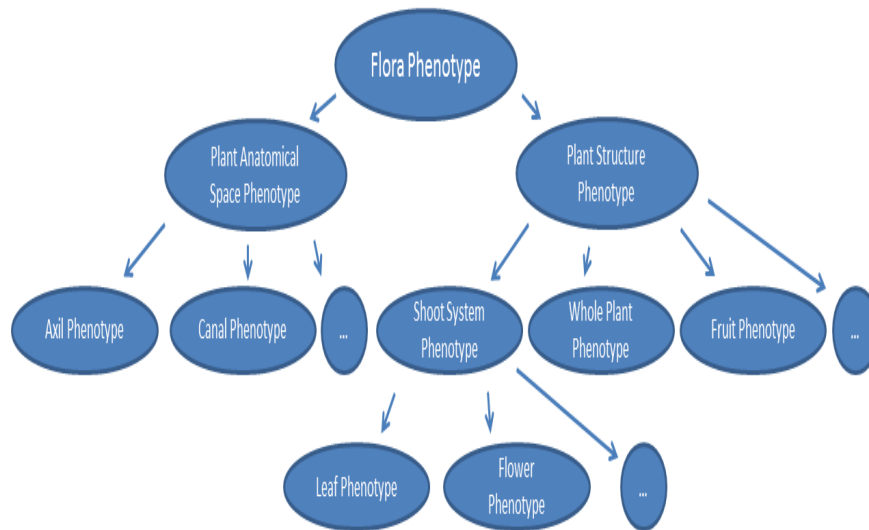
**Description:** Viscid perennial with very pale lemon-yellow flowers

**Collector:** M.F. Gardner & S.G. Knees No. 9308

**Date:** 31st October 2015

# Flora Phenotype Ontology (FLOPO)

A domain ontology and knowledge base of flowering plant traits (Hoehndorf et al. 2016).



species_name	trait_name
"Adansonia digitata L."@en	"leaf alternate placement"^^<http://www.w3.org/2001/XMLSchema#st
"Adansonia digitata L."@en	"fruit hairy"^^<http://www.w3.org/2001/XMLSchema#st
"Adansonia digitata L."@en	"leaf simple"^^<http://www.w3.org/2001/XMLSchema#st
"Adansonia digitata L."@en	"petal amount 5"@en
"Adansonia digitata L."@en	"flower white"^^<http://www.w3.org/2001/XMLSchema#st
"Adansonia digitata L."@en	"inflorescence type solitary"@en
"Adansonia digitata L."@en	"tissue succulent"@en
"Adansonia digitata L."@en	"leaf digitate"^^<http://www.w3.org/2001/XMLSchema#st
"Adansonia digitata L."@en	"shoot apex ovate"^^<http://www.w3.org/2001/XMLSchema#st
"Adansonia digitata L."@en	"filament glabrous"^^<http://www.w3.org/2001/XMLSchema#st
"Adansonia digitata L."@en	"bud ovate"^^<http://www.w3.org/2001/XMLSchema#st
"Adansonia digitata L."@en	"shoot apex spherical"^^<http://www.w3.org/2001/XMLSchema#st
"Adansonia digitata L."@en	"bark grey"^^<http://www.w3.org/2001/XMLSchema#st
"Adansonia digitata L."@en	"stem conical"^^<http://www.w3.org/2001/XMLSchema#st
"Adansonia digitata L."@en	"shoot system pubescent"^^<http://www.w3.org/2001/XMLSchema#st
"Adansonia digitata L."@en	"crown root conical"^^<http://www.w3.org/2001/XMLSchema#st
"Adansonia digitata L."@en	"bud pointed"^^<http://www.w3.org/2001/XMLSchema#st
"Adansonia digitata L."@en	"shoot apex pointed"^^<http://www.w3.org/2001/XMLSchema#st
"Adansonia digitata L."@en	"stigma star shaped"^^<http://www.w3.org/2001/XMLSchema#st
"Adansonia digitata L."@en	"whole plant phanerophytic"@en
"Adansonia digitata L."@en	"flower actinomorphic"^^<http://www.w3.org/2001/XMLSchema#st
"Adansonia digitata L."@en	"fruit indehiscent"@en

# Plant Traits

Sample traits for leaves and flowers (Dressler et al. 2014).

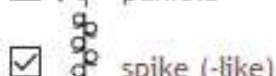
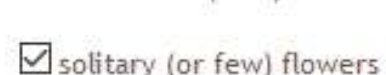
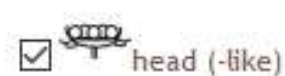
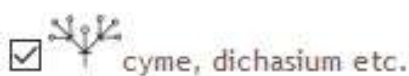
## Venation



## Arrangement



## Inflorescence



## Image-Trait Dataset

- Total herbarium scans: 13157
- Training: 70%, Validation: 10%, Test: 20%
- Total species: 2339
- Selected plant traits: 27
- Leaf Traits: 14, Flower Traits: 9, Fruit Traits: 4

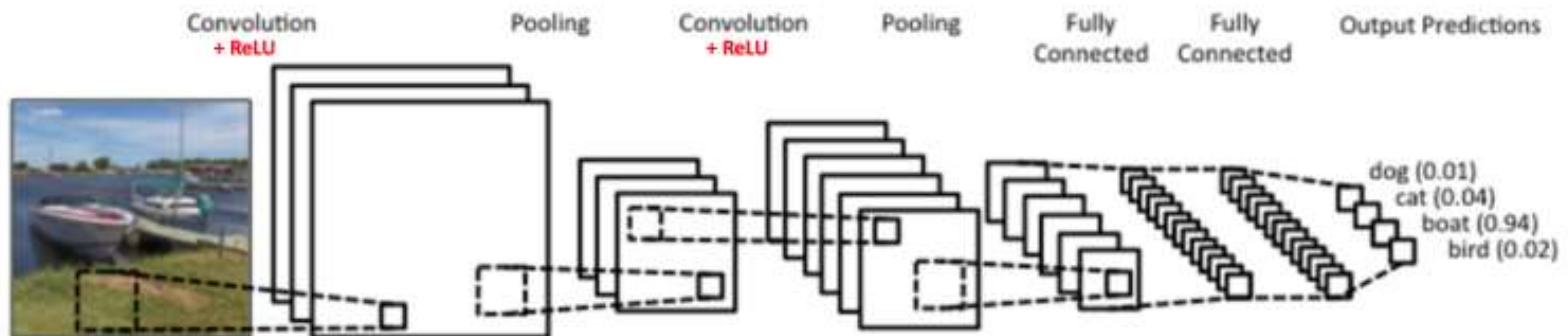
## Image-Trait Dataset

Limiting factors for number of images:

- Very low percentage of herbarium records with digitized collector notes.
- Sparse occurrence of flower and fruit key words in many collector notes.
- Relatively low number of species of herbarium represented in FLOPO due to differing geographical focus and collector bias.

# Deep Learning Approach

Feature extraction by deep convolutional neural network.



A simple convolutional neural network. (clarifai.com)

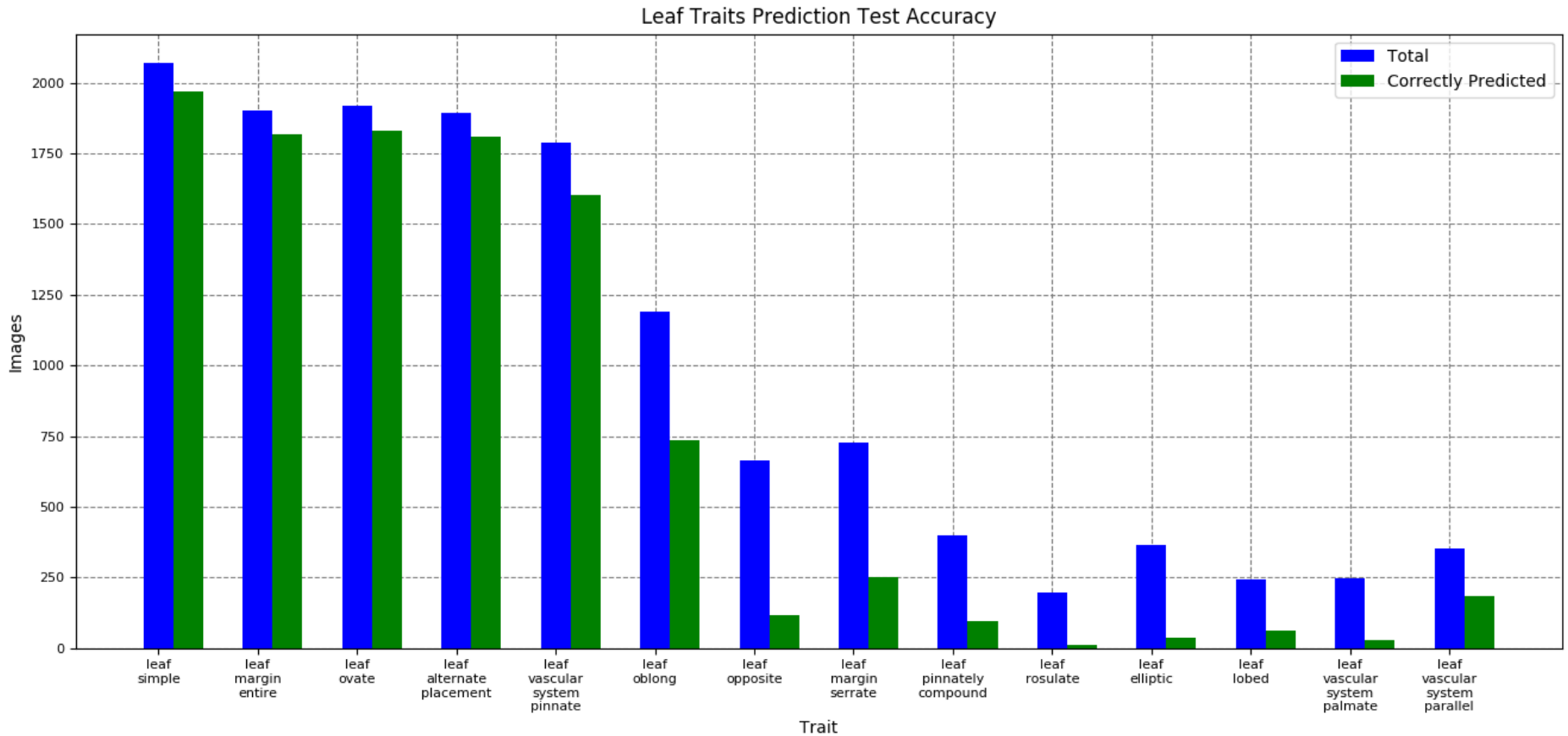
# Deep Learning Approach

## Modified ResNet18 (He et al. 2016)

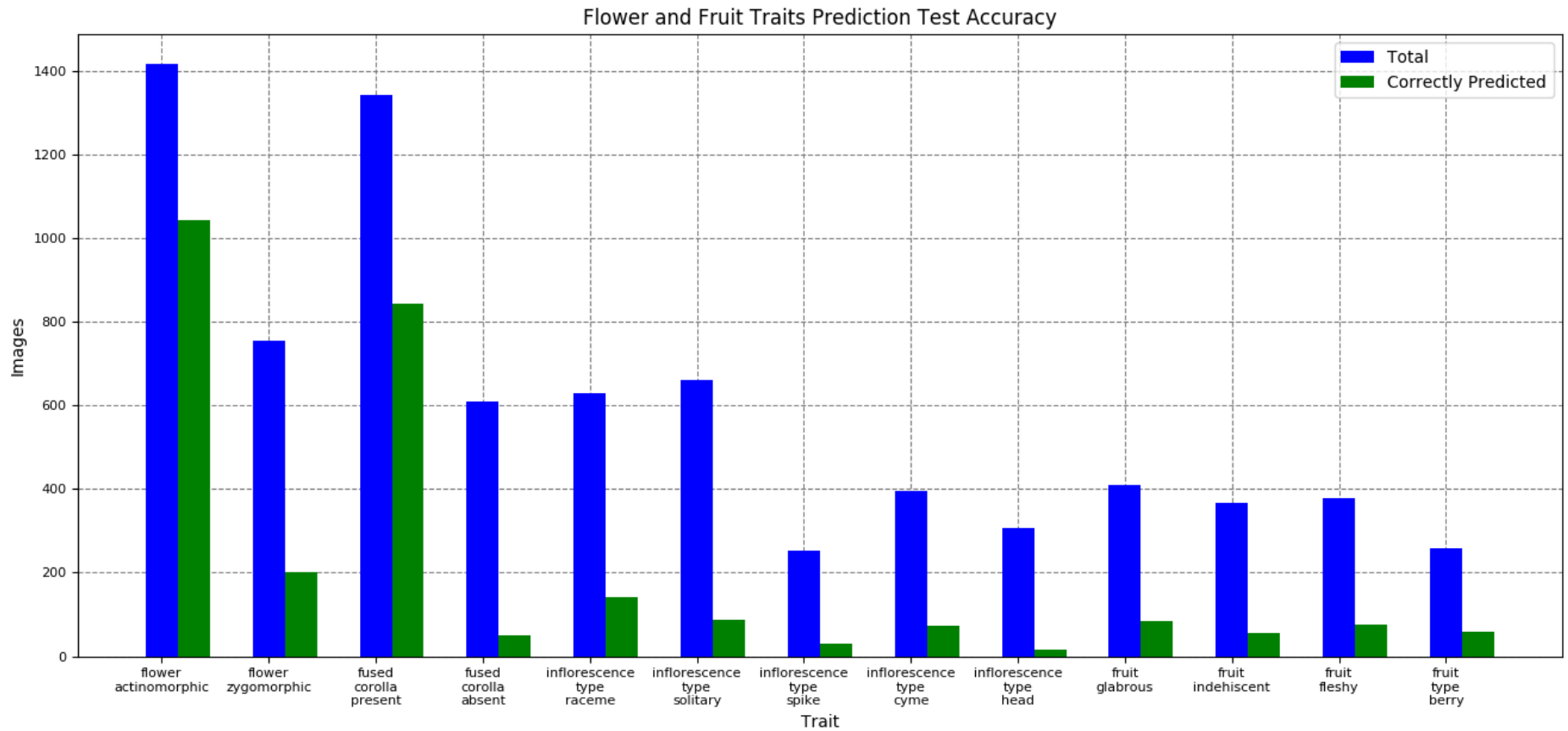
Layer Type	Filter Size / Stride / Padding	Output Size (C x H x W)
Input		3 x 1200 x 800
Convolution	7 x 7 / 2 / 3	32 x 600 x 400
Max Pool	3 x 3 / 2 / 1	32 x 300 x 200
Convolution	7 x 7 / 2 / 3	64 x 150 x 100
Max Pool	3 x 3 / 2 / 1	64 x 75 x 50
Convolution	[3 x 3, 3 x 3] x 2	64 x 75 x 50
Convolution	1 x 1 / 2, [3 x 3, 3 x 3] x 2	128 x 38 x 25
Convolution	1 x 1 / 2, [3 x 3, 3 x 3] x 2	256 x 19 x 13
Convolution	1 x 1 / 2, [3 x 3, 3 x 3] x 2	512 x 10 x 7
Average Pool	10 x 7 / 10 x 7	512 x 1 x 1
Fully Connected	27 dense	27



# Result

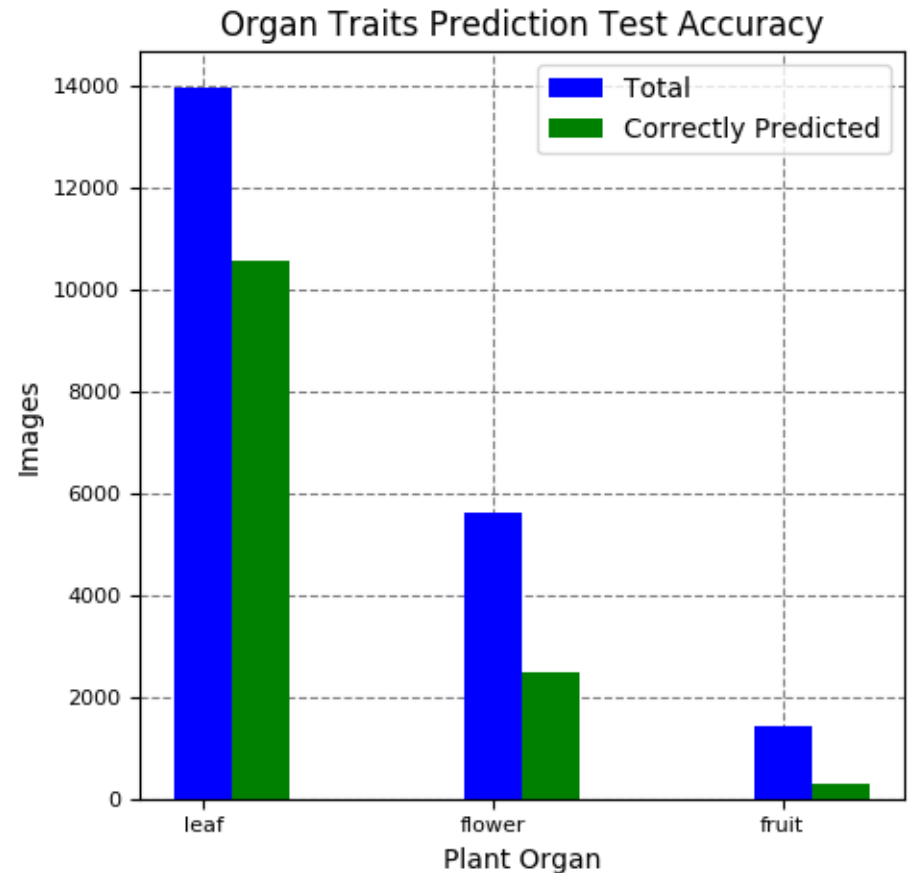


# Result



## Discussion

- Sampling bias for plant organs.
- More accurate traits have more samples.
- Traits for flowers and fruits have low accuracy.



## Outlook

- More herbarium / plant collections with digitized / computer readable labels or annotations.
- Plant images with standardized annotated traits and features from TRY and other datasets.
- Identify region / heatmap of image contributing to each trait (Samek et al. 2017).
- Create application to assist and automate trait and taxon recognition / detection on plant images.

## Thanks to

- DFG for funding this project
- Herbarium collectors and data providers
- Deep learning community

## Selected FLOPO Traits

OBO ID	Trait	OBO ID	Trait
FLOPO:0000693	Leaf simple	FLOPO:0900002	Inflorescence type raceme
FLOPO:0900073	Leaf margin entire	FLOPO:0900007	Inflorescence type solitary
FLOPO:0000286	Leaf ovate	FLOPO:0900001	Inflorescence type spike
FLOPO:0001032	Leaf alternate placement	FLOPO:0900006	Inflorescence type cyme
FLOPO:0000103	Leaf oblong	FLOPO:0900005	Inflorescence type head
FLOPO:0000420	Leaf opposite	FLOPO:0001316	Flower actinomorphic
FLOPO:0900074	Leaf margin serrate	FLOPO:0005591	Flower zygomorphic
FLOPO:0907004	Leaf pinnately compound	FLOPO:0907002	Fused corolla present
FLOPO:0900066	Leaf rosulate	FLOPO:0907003	Fused corolla absent
FLOPO:0000579	Leaf elliptic	FLOPO:0000600	Fruit glabrous
FLOPO:0006834	Leaf lobed	FLOPO:0900020	Fruit indehiscent
FLOPO:0000561	Leaf vascular system pinnate	FLOPO:0002102	Fruit fleshy
FLOPO:0900070	Leaf vascular system palmate	FLOPO:0900015	Fruit type berry
FLOPO:0900072	Leaf vascular system parallel		

## References

- Dressler, Stefan, Marco Schmidt, and Georg Zizka. "Introducing African Plants—A Photo Guide—An Interactive Photo Data-Base and Rapid Identification Tool for Continental Africa." *Taxon* 63.5 (2014): 1159-1161.
- He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- Hoehndorf, Robert, et al. "The flora phenotype ontology (FLOPO): tool for integrating morphological traits and phenotypes of vascular plants." *Journal of biomedical semantics* 7.1 (2016): 65.
- Pound, Michael P., et al. "Deep machine learning provides state-of-the-art performance in image-based plant phenotyping." *GigaScience* (2017).
- Samek, Wojciech, et al. "Evaluating the visualization of what a deep neural network has learned." *IEEE transactions on neural networks and learning systems* 28.11 (2017): 2660-2673.
- Ubbens, Jordan R., and Ian Stavness. "Deep plant phenomics: a deep learning platform for complex plant phenotyping tasks." *Frontiers in plant science* 8 (2017): 1190.
- Younis, Sohaib, et al. "Taxon and trait recognition from digitized herbarium specimens using deep convolutional neural networks." *Botany Letters* (2018): 1-7.