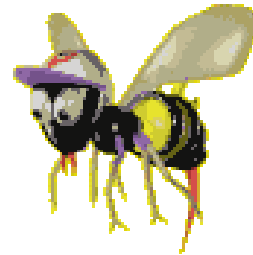


Accelerating taxonomic research using **image recognition systems** and **citizen-science contributions**



Phd Student: Miroslav Valan
Supervisor: Fredrik Ronquist
Co-supervisors: Atsuto Maki, Karoly Makonyi

Conclusion

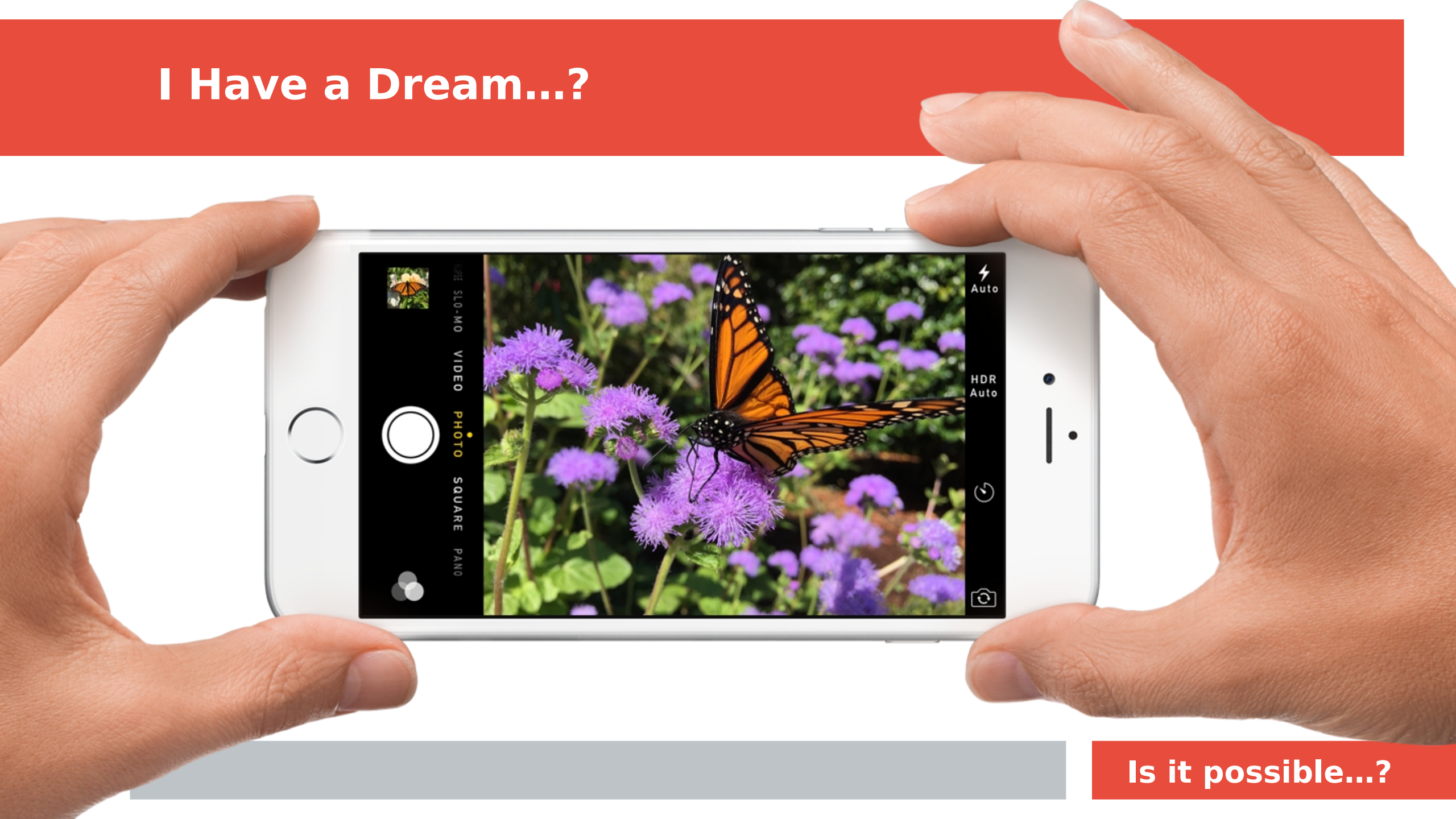
Machines don't bite...
see what we cannot....
learn much faster
never forget



Message...

Collect machine readable data
images, videos...
sounds,...

I Have a Dream...?

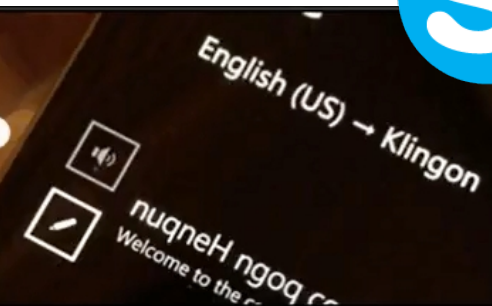
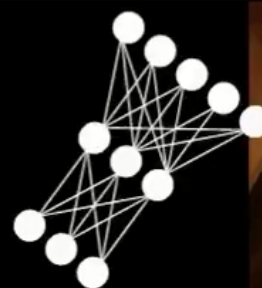


Is it possible...?

Some machine learning tasks



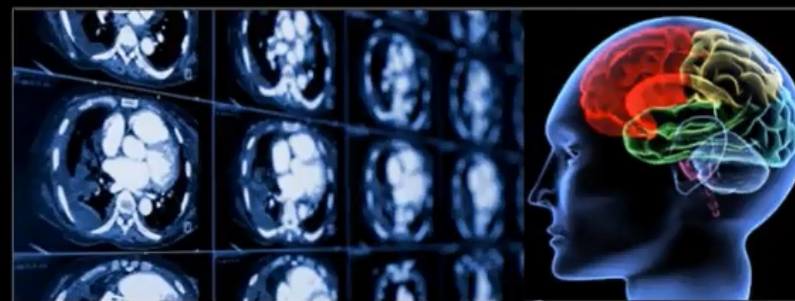
Image Classification, Object Detection, Localization, Action Recognition, Scene Understanding



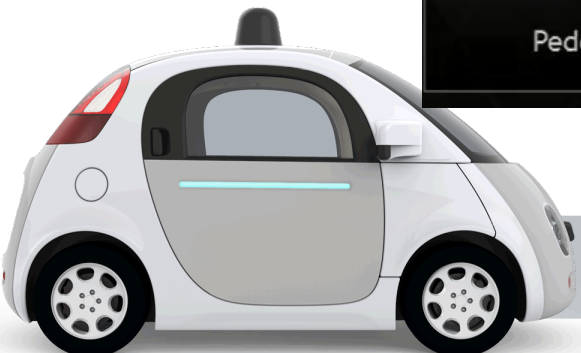
Speech Recognition, Speech Translation, Natural Language Processing



Pedestrian Detection, Traffic Sign Recognition



Breast Cancer Cell Mitosis Detection, Volumetric Brain Image Segmentation



Deep learning & AI

powerful learning
algorithms
(DEEP LEARNING)

FREE

sufficient computational
power (GPU)

training data
(images, videos,...)



small dataset < petabyte (1000 TB)

256 x 256 ~ 10-50 KB

<https://developer.nvidia.com/deep-learning-courses>

lightning

deformation

scale

**direction
of the object**

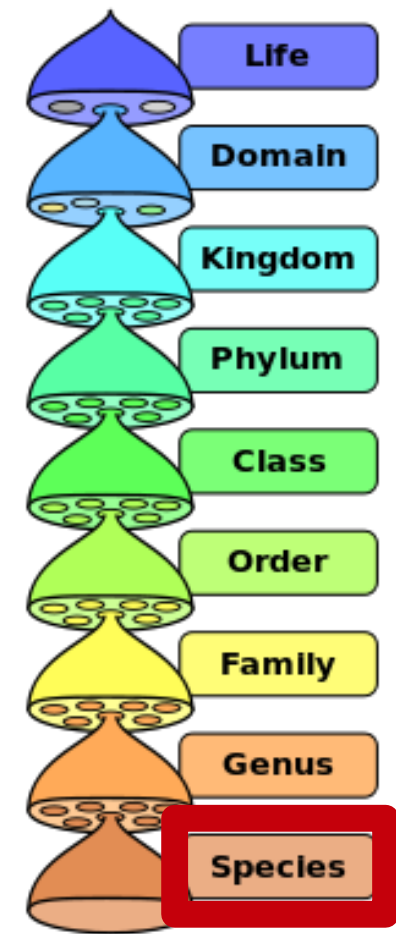
viewpoint

**2D
representation of a 3D
object**

**intra-class
variation**

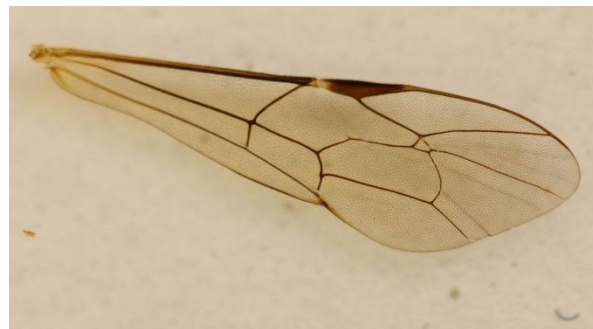


Present automated taxon identification (ATI)



Wing advantages

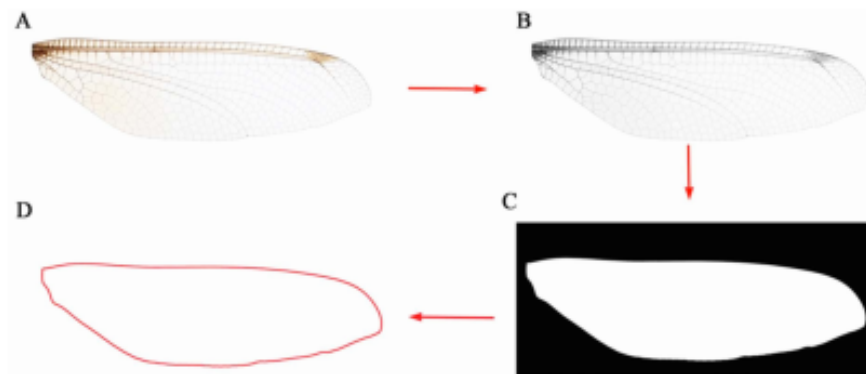
- flat - almost 2D
- imaging settings:
 - viewpoint
 - object orientation
 - lightning



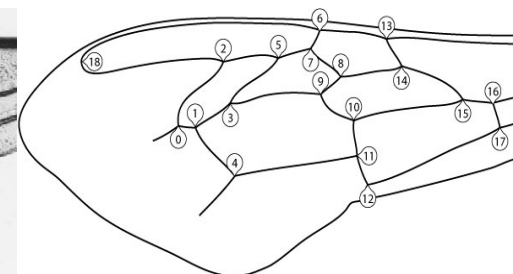
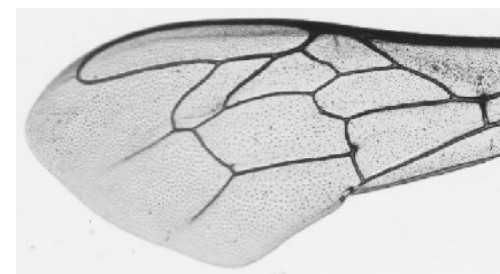
Disadvantages

- time consuming
- destructive
- hard to elicit public interest
- template not a pattern

DAIIS



DrawWing



Hand-crafted features (human involved)

Tofilski 2008, Yang 2015



Cryptini



Cyloceria



Hemigastrini



Hemigastrini



Campopleginae



Campopleginae



Pimplini



Campopleginae



Hemigastrini



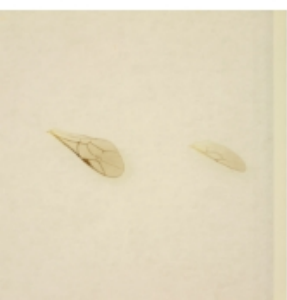
Hemigastrini



Campopleginae



Hemigastrini



Hemigastrini



Pimplini



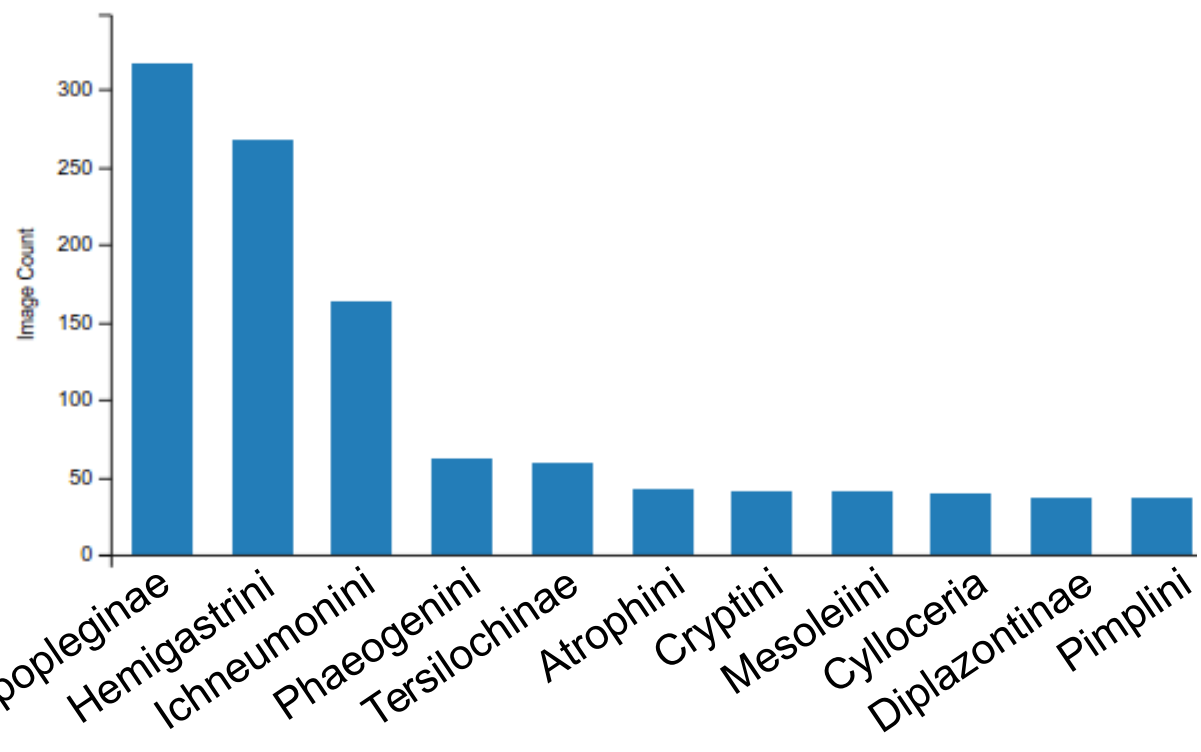
Campopleginae



Hemigastrini

Ichneumonidae - subfamilies and tribes

Number of Categories	11	
Training (on chart)	1037	(70.0%)
Validation	326	(22.0%)
Test	123	(8.0%)





Cryptini



Cyloceria



Hemigastrini



Hemigastrini



Campopleginae



Campopleginae



Pimplini



Campopleginae



Hemigastrini



Hemigastrini



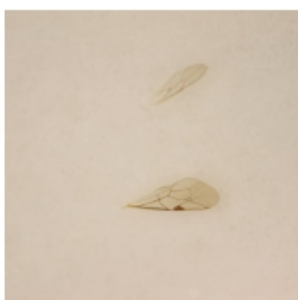
Campopleginae



Hemigastrini



Hemigastrini



Pimplini



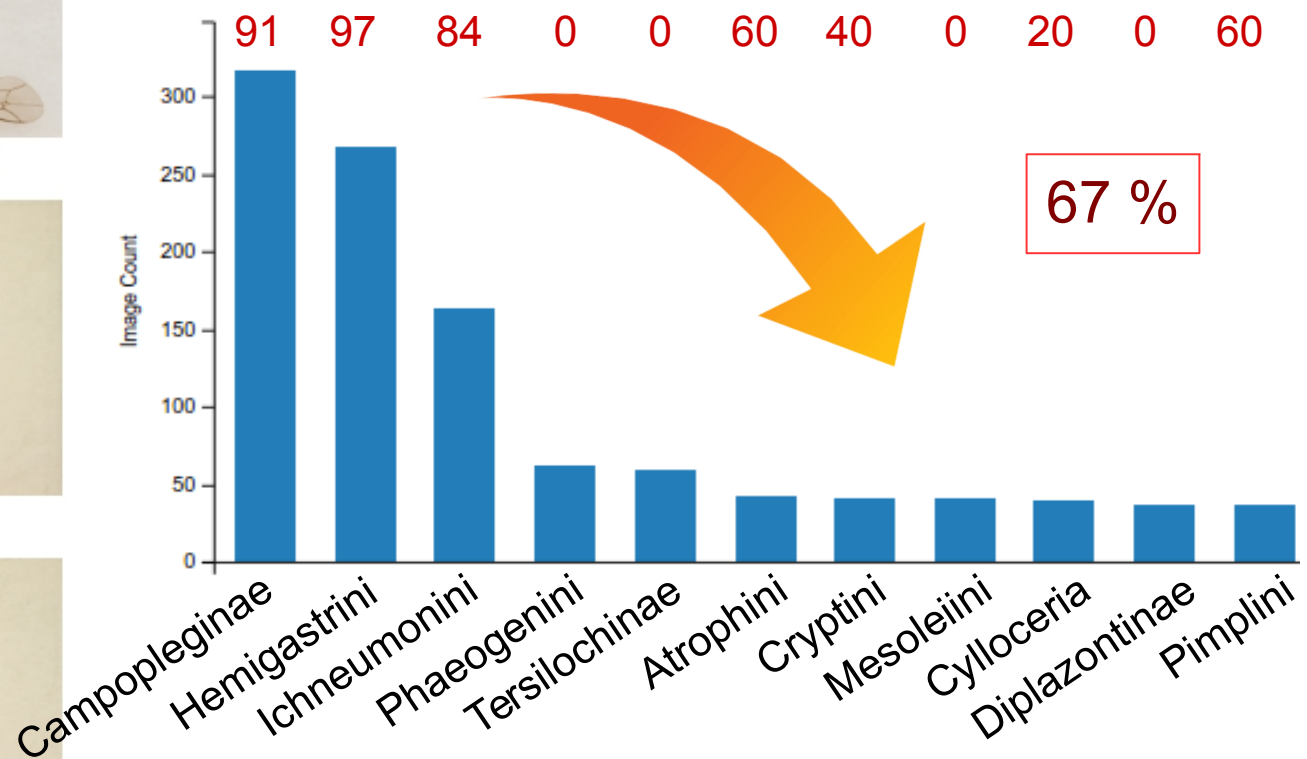
Campopleginae



Hemigastrini

Ichneumonidae - subfamilies and tribes

Number of Categories	11	
Training (on chart)	1037	(70.0%)
Validation	326	(22.0%)
Test	123	(8.0%)





Tabanidae



Asilidae



Asilidae



Tabanidae

Diptera's face recognition- family level

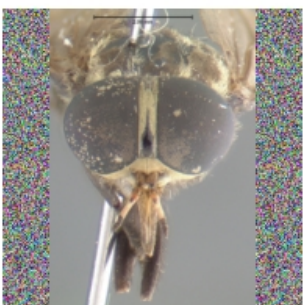
Number of Categories	11
Training (on chart)	672 (75.0%)
Validation	179 (20.0%)
Test	43 (5.0%)



Asilidae



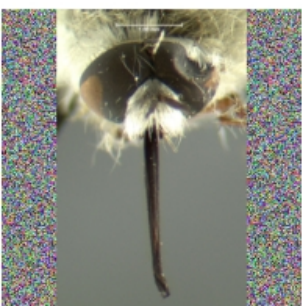
Empididae



Tabanidae



Asilidae



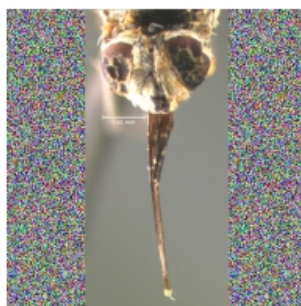
Bombyliidae



Bombyliidae



Syrphidae



Bombyliidae



Tabanidae



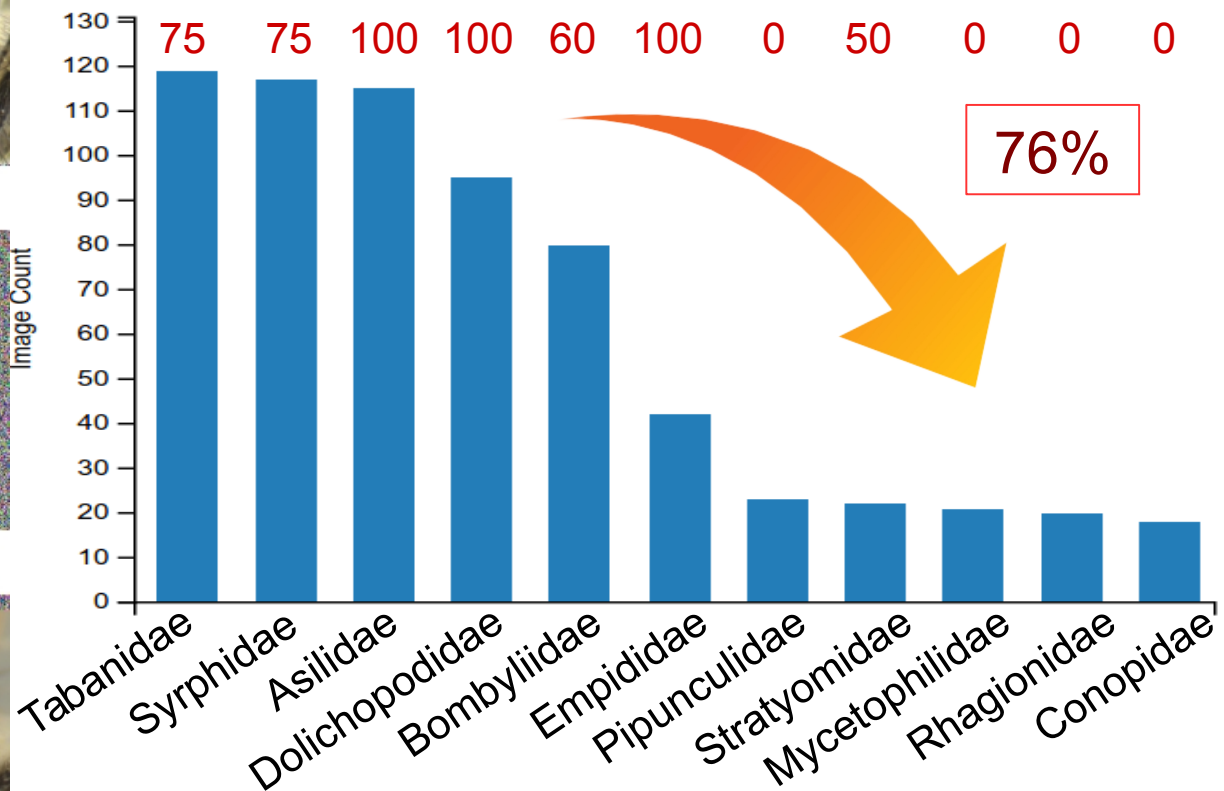
Tabanidae



Syrphidae



Asilidae



Diptera's face recognition- genus level

Number of Categories	6
Training (on chart)	162 (75.0%)
Validation	43 (20.0%)
Test	10 (5.0%)



Tabanus Tabanidae



Tabanus Tabanidae



Dolichopus Dolichopodidae



Tabanus Tabanidae



Anthrax Bombyliidae



Dolichopus Dolichopodidae



Cyrtopogon Asilidae



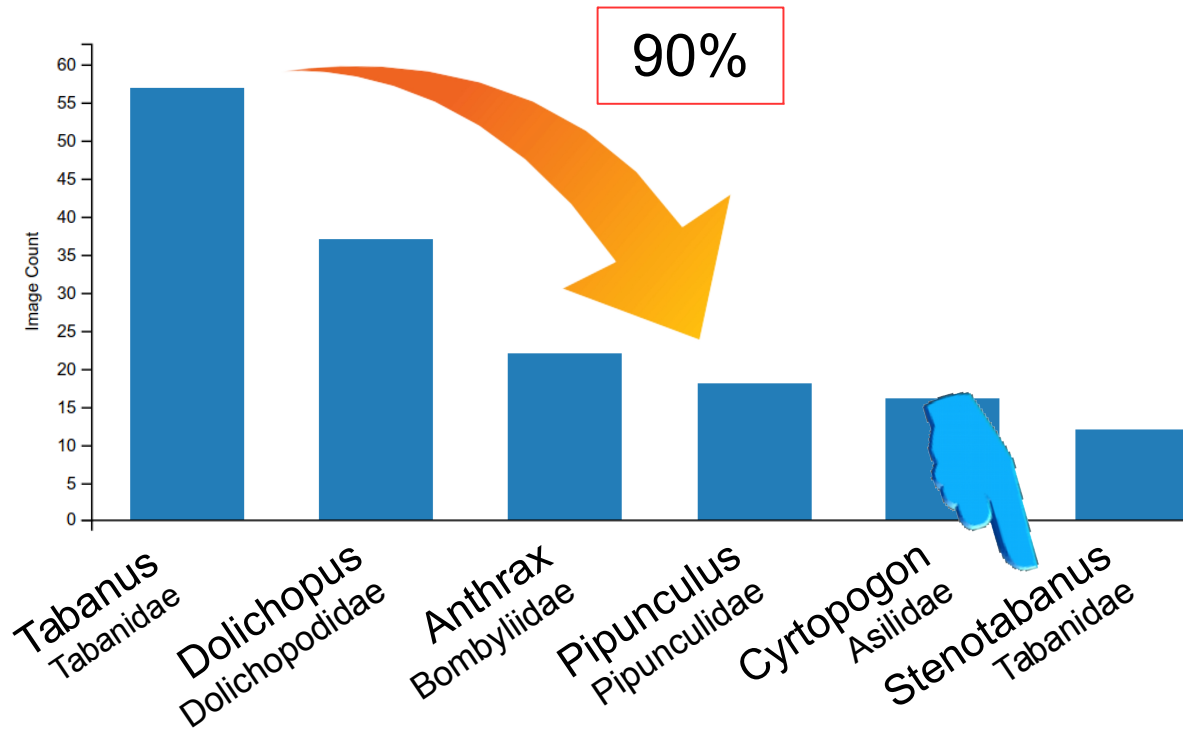
Stenotabanus Tabanidae

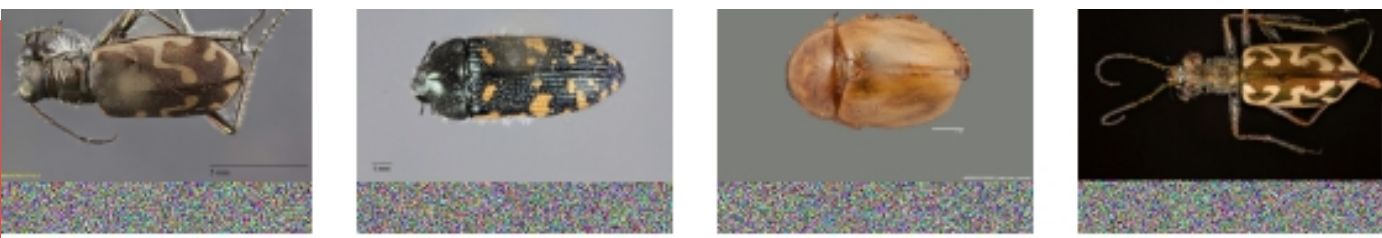


Tabanus Tabanidae



Pipunculus Pipunculidae





Coleoptera top view - family level

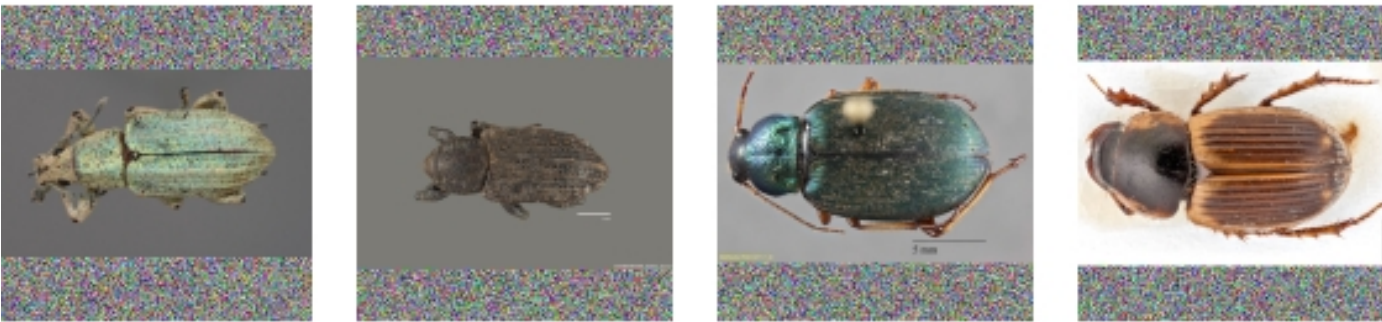
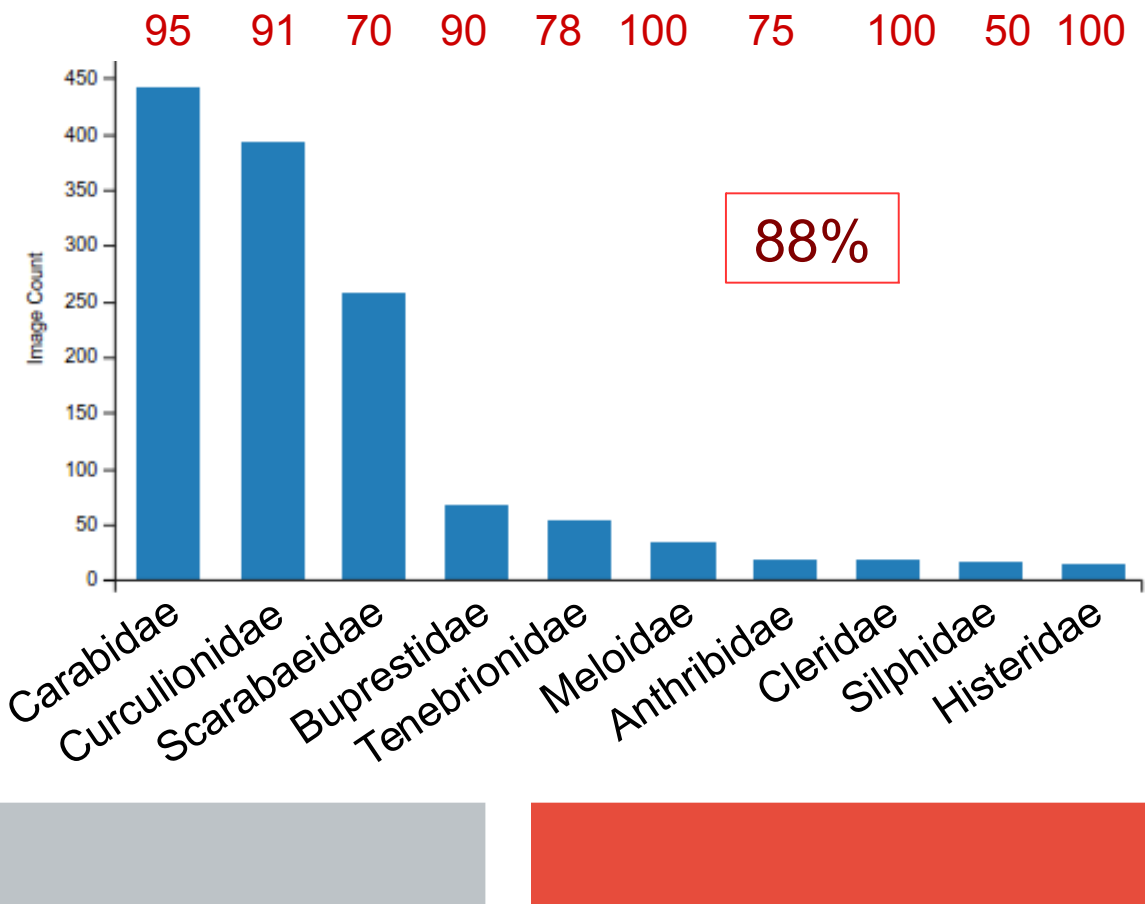
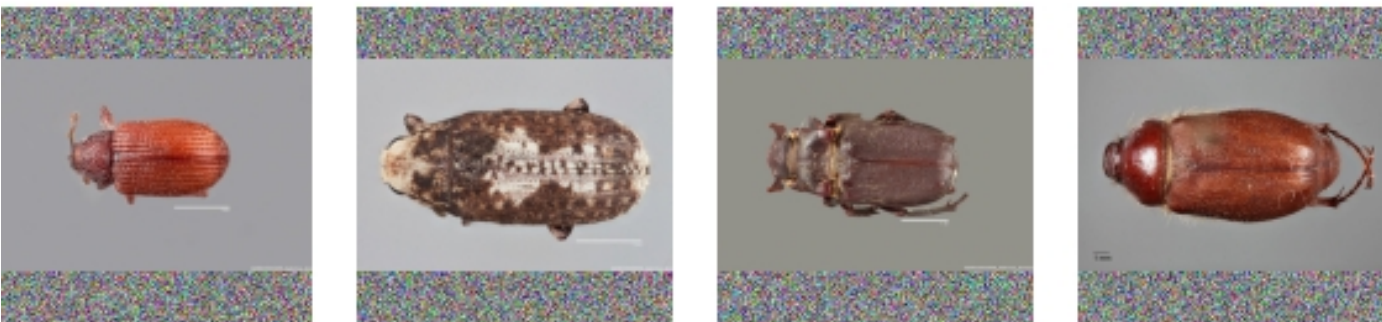
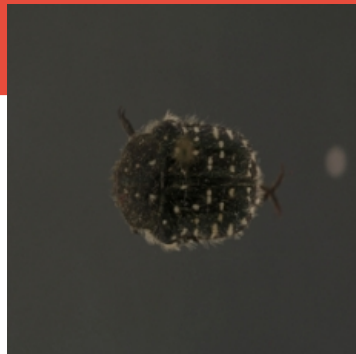


Table with 3 columns: Category, Count, and Percentage. Rows include Number of Categories (10), Training (on chart) (1312, 65.0%), Validation (485, 25.0%), and Test (201, 10.0%).



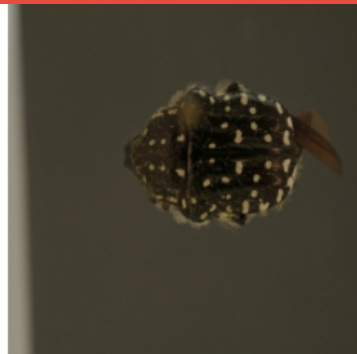
Coleoptera top view - three species



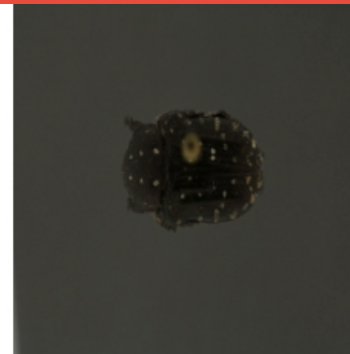
Oxythyrea funesta



Oxythyrea subcalva



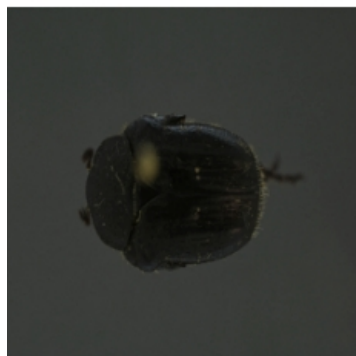
Oxythyrea pantherina



Oxythyrea funesta

340 specimens

Training	(60.0%)
Validation	(30.0%)
Test	(10.0%)



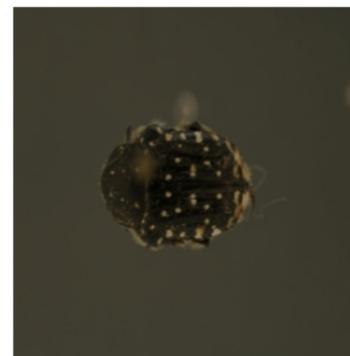
Oxythyrea subcalva



Oxythyrea pantherina



Oxythyrea pantherina



Oxythyrea pantherina

100%

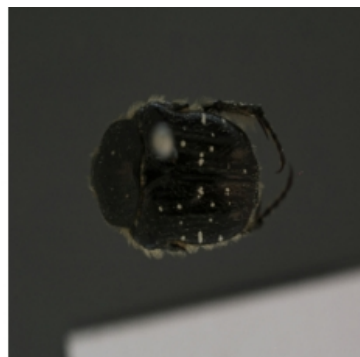
Oxythyrea funesta

Oxythyrea pantherina

Oxythyrea subcalva



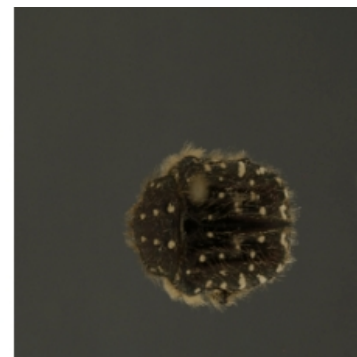
Oxythyrea pantherina



Oxythyrea subcalva



Oxythyrea funesta



Oxythyrea funesta

Images taken and labeled by
Dominik Vondráček

Bugs bug the bug master but bugs matter to the bug master



miroslav.valan@nrm.se
miroslav@savanticab.com



This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 642241

Accelerating taxonomic research using
image recognition systems
and citizen-science contributions



Phd Student: Miroslav Valan
Supervisor: Fredrik Ronquist
Co-supervisors: Atsuto Maki, Karoly Makonyi

Hej,
I am Miroslav and I'll shortly introduce my project.

Please note the three key words : taxonomy, image
recognition and citizen-science

Conclusion

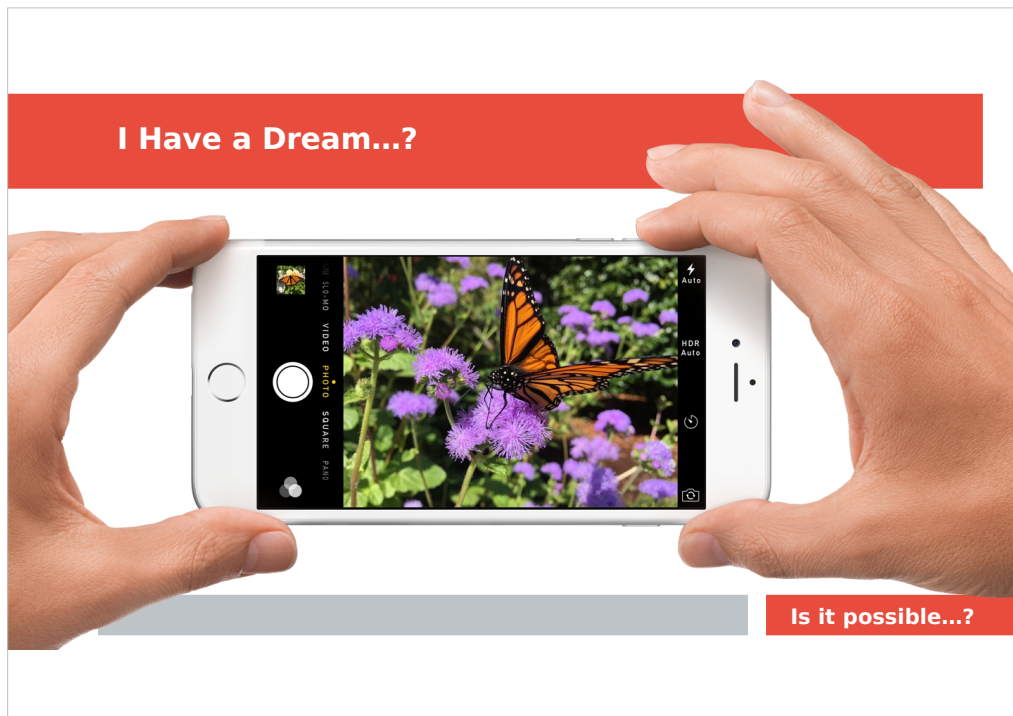
Machines don't bite...
see what we cannot....
learn much faster
never forget



Message...

Collect machine readable data
images, videos...
sounds,...



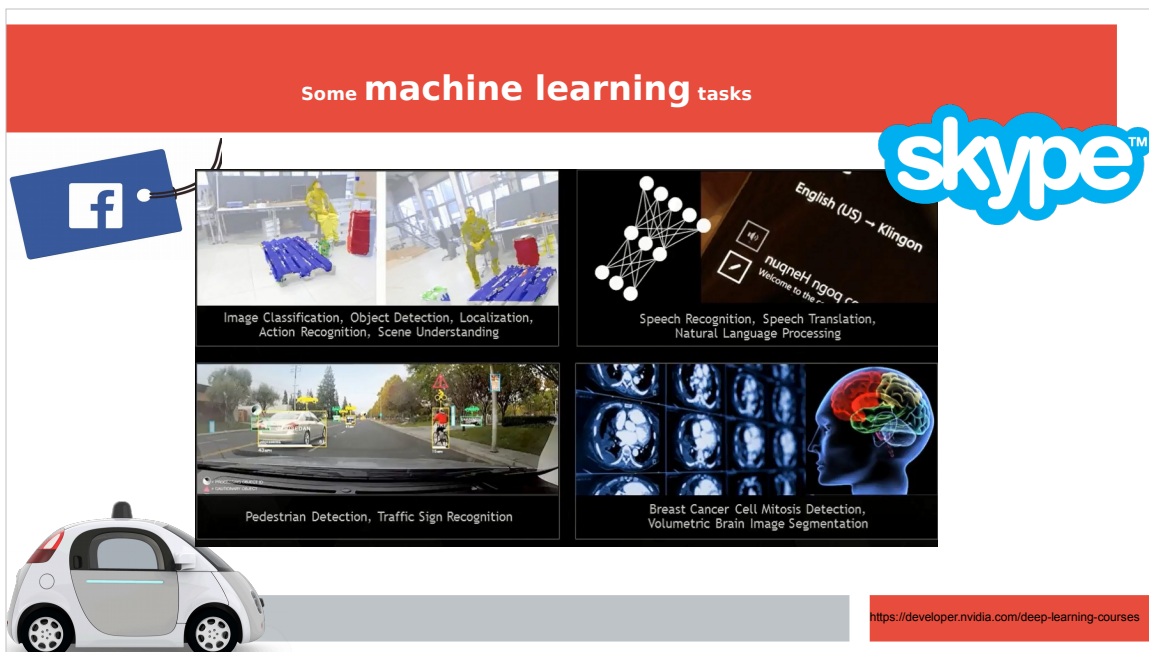


Imagine being able to identify insects by just photographing them with your smartphone!

It could give you the right ID in a second and link you to more sources on Google

When image is uploaded locality is stored and occurrence is linked to GBIF.

Scientists with special interest in taxon could be instantly notified.



We are particularly interested in the upper left examples
Computer Vision

Machines are inspired by the way the human eye and
brain work.

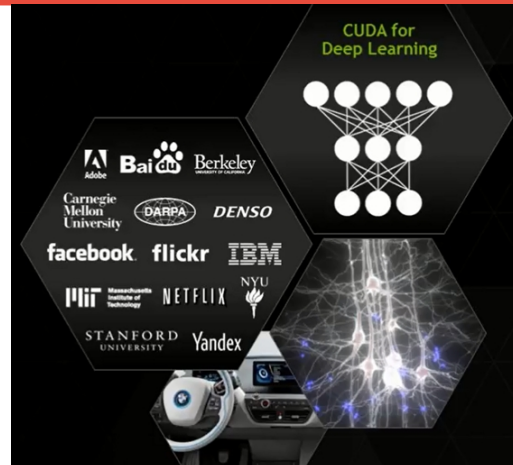
Deep learning & AI

powerful learning
algorithms
(DEEP LEARNING)

FREE

sufficient computational
power (GPU)

training data
(images, videos,...)



small dataset < petabyte (1000 TB)

256 x 256 ~ 10-50 KB

<https://developer.nvidia.com/deep-learning-courses>

It used to be that when you wanted to get computer do something new you would have to program it.

But not anymore, now machines could learn just by themselves how to be better and it is called machine learning.

Machines will help us in moving "the dream" closer to reality. Machines powered by deep -learning algorithms and GPU-s perceive and understand the world same as we do.

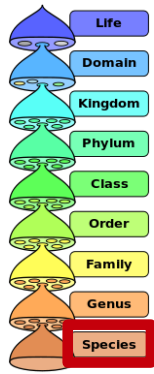
It has been successfully used for production and for research.

Sometimes it is even creepy, when facebook or linkedin suggest who your friends might be and you dont have idea how it did it. They used machine learning. Algorithms that can learn from data rather than being programmed by hand.

Even more amazing is driverless Google car.

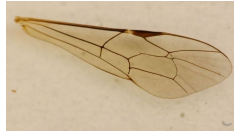


Present automated taxon identification (ATI)



Wing advantages

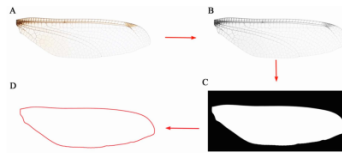
- flat - almost 2D
- imaging settings:
 - viewpoint
 - object orientation
 - lightning



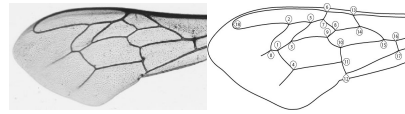
Disadvantages

- time consuming
- destructive
- hard to elicit public interest
- template not a pattern

DAIIS

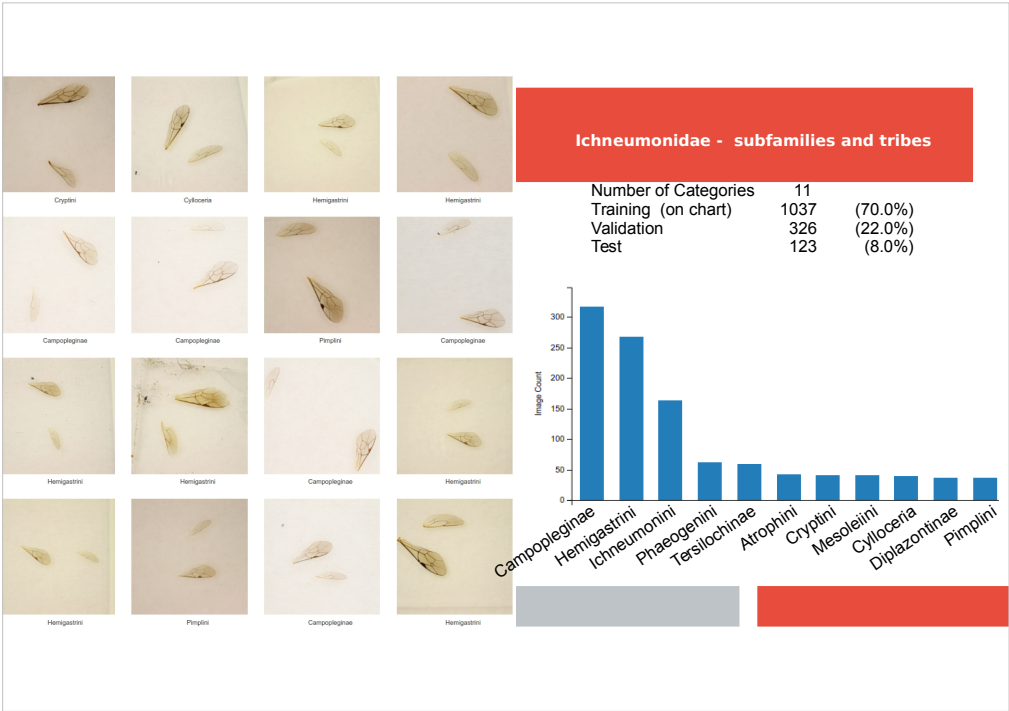


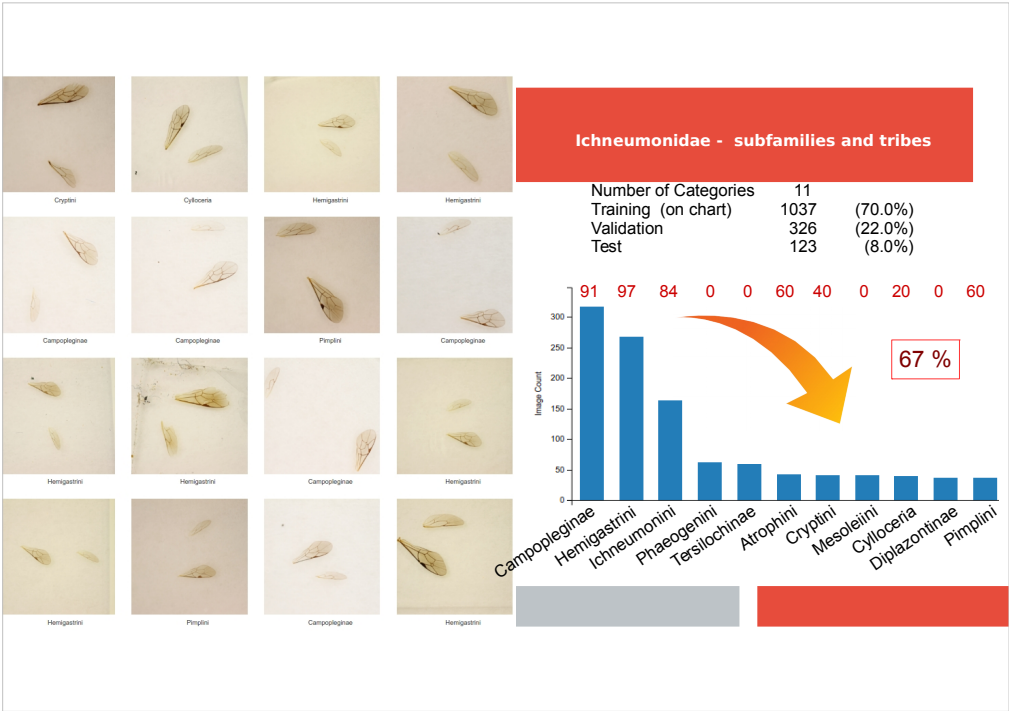
DrawWing

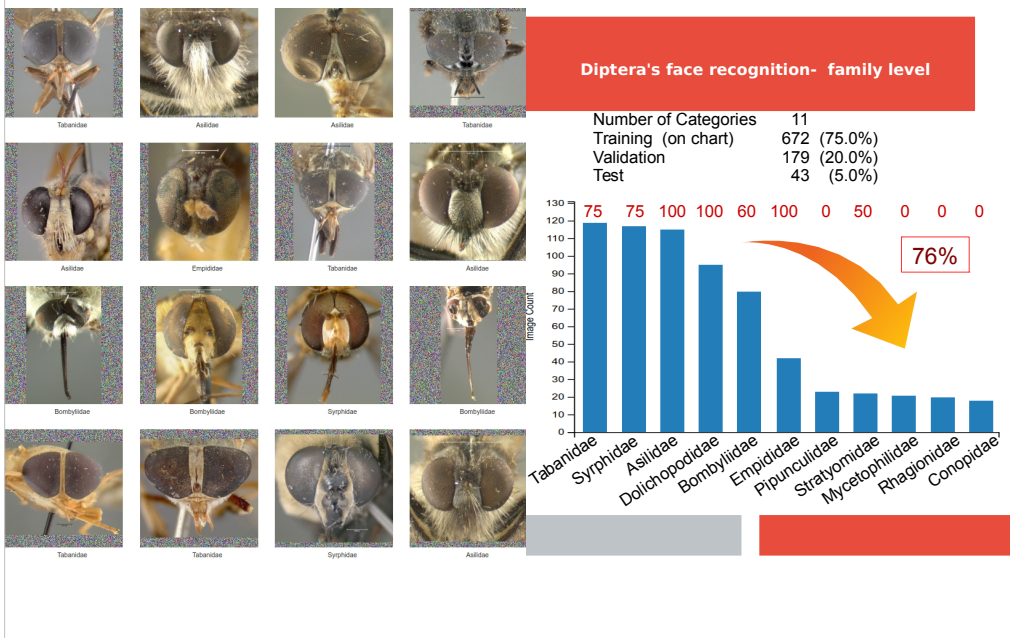


Hand-crafted features (human involved)

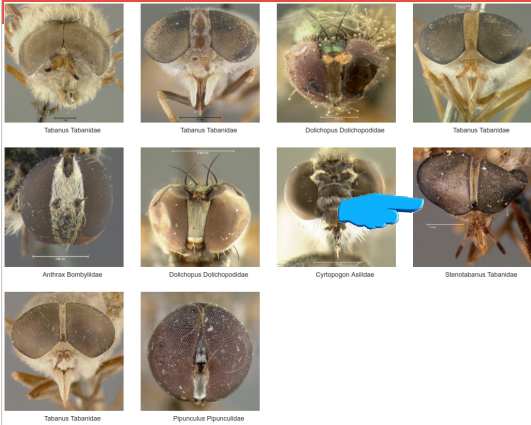
Tofliski 2008, Yang 2015



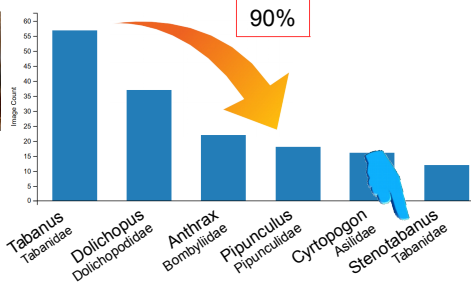


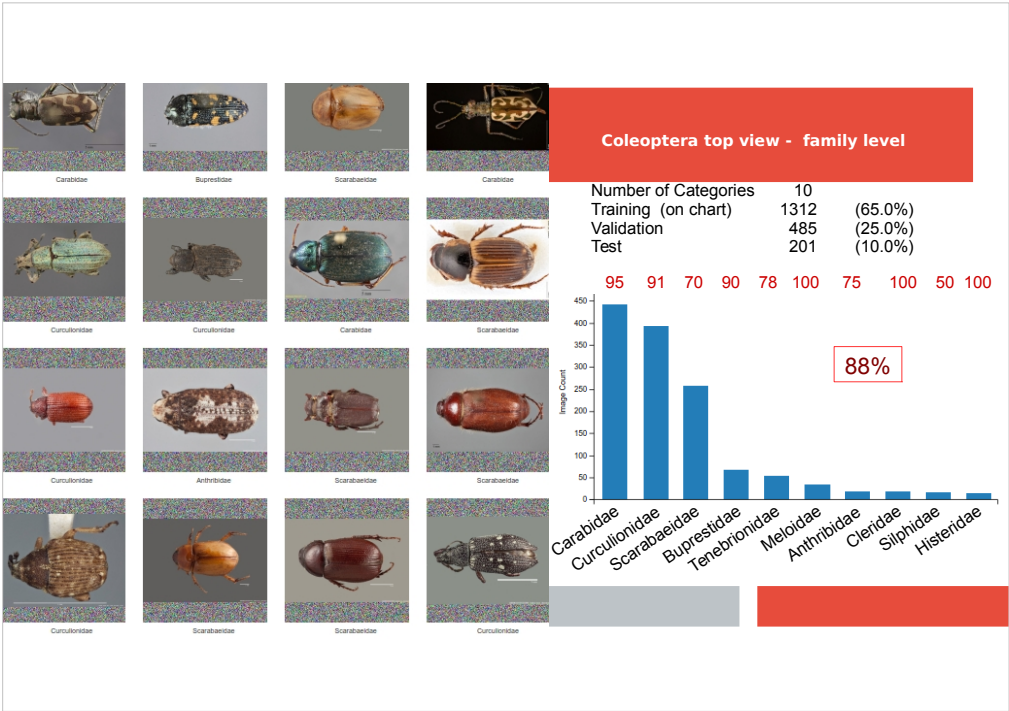


Diptera's face recognition- genus level

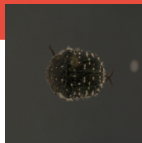


Number of Categories	6
Training (on chart)	162 (75.0%)
Validation	43 (20.0%)
Test	10 (5.0%)





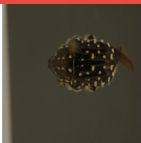
Coleoptera top view - three species



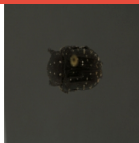
Oxythyrea funesta



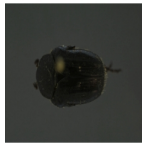
Oxythyrea subcalva



Oxythyrea pantherina



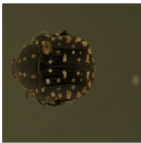
Oxythyrea funesta



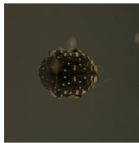
Oxythyrea subcalva



Oxythyrea pantherina



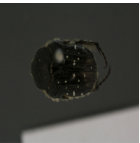
Oxythyrea pantherina



Oxythyrea pantherina



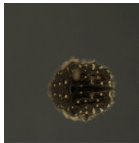
Oxythyrea pantherina



Oxythyrea subcalva



Oxythyrea funesta



Oxythyrea funesta

340 specimens

Training (60.0%)
Validation (30.0%)
Test (10.0%)

100%

Oxythyrea funesta
Oxythyrea pantherina
Oxythyrea subcalva

Images taken and labeled by
Dominik Vondráček

Bugs bug the bug master but bugs matter to the bug master



miroslav.valan@nrm.se
miroslav@savanticab.com



This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 642241