

High throughput phenotyping for the phenomics era

Moritz Lürig^{*1}, Blake Matthews^{*}

^{*} EAWAG, Department of Aquatic Ecology; ¹ ETH Zürich, Center for Adaptation to a Changing Environment (ACE)

Scientific process

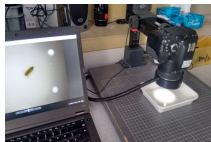
Experiments



Organisms



Image acquisition



Problem description

It is widely acknowledged that the integration of molecular and phenotypic information holds the key to understanding many important biological phenomena. However, while most molecular methods can achieve high throughput ("omics"), the collection and processing of phenotypic data is still a mostly manual task and thus often painstakingly slow.

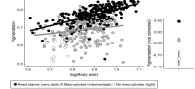
Here we present the python package **phenotype**, which builds on other scientific python modules (opencv, numpy, pandas,...), creating an integrated environment to process digital images [1] and segment them [2] to extract the desired information [3].

phenotype is available at <https://github.com/mluerig/phenotype>

Scientific integration



Phenotypic information



Post-processing and statistics



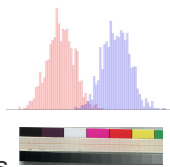
phenotype

a phenotyping pipeline
for python

[1] Pre-processing

Edit image meta-data to make projects comparable and data reproducible.

- Adjust exposure, brightness and colour
- Measure scale and calculate pixel ratios
- A central project file keeps track of this information and links it to each image



[2] Segmentation

Find parts of image that are labels, organisms, arena – core function.

- Otsu- and adaptive binarization

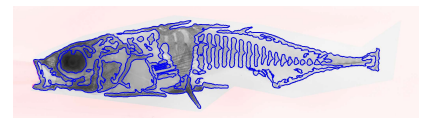


- Movement detection (foreground-background subtraction)
- OCR recognition for text using **pytesseract**

[3] Feature extraction

Save various information to textfiles for further processing and statistics

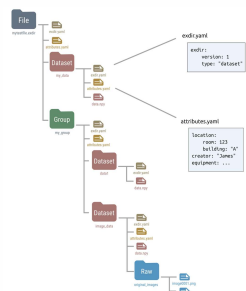
- Pixel values (HSV, grayscale)



- Morphology (area, shape, landmarks)
- Movement (trajectories, calculate using **trackpy**)
- Detection of scale and labels

Outlook

Improve data management, e.g. with experimental directory structure (**Exdir**)



Use images to create training data for neural networks, e.g. **Mask R-CNN**



Conclusions

The need for flexibility in the collection of phenotypic data from digital images has led to the development of **phenotype**. The python programming language with its scientific modules, as well as an extremely active online community has proven extremely useful in the development of this tool. **phenotype** is in alpha stage, but will be developed further to improve user friendliness while maintaining versatility, which we believe is its key asset.

List of cited python packages (non-standard libraries):

Exdir - Directory structure standard for experimental pipelines <https://github.com/CINPLA/exdir>
Mask R-CNN - object detection and segmentation https://github.com/matterport/Mask_RCNN
Opencv - Open Source Computer Vision Library <https://pyip.org/project/opencv-contrib-python/>
Pytesseract - wrapper for Google's Tesseract-OCR <https://pypi.org/project/pytesseract/>
Trackpy - Fast, Flexible Particle-Tracking Toolkit <http://soft-matter.github.io/trackpy/v0.4.1/>

Moritz Lürig | luerig.net
 Twitter / github | [mluerig](https://github.com/mluerig)
 Doctoral student | ETH Zürich / Eawag
 Supervisors | Blake Matthews, Jukka Jokela



eawag
 aquatic research