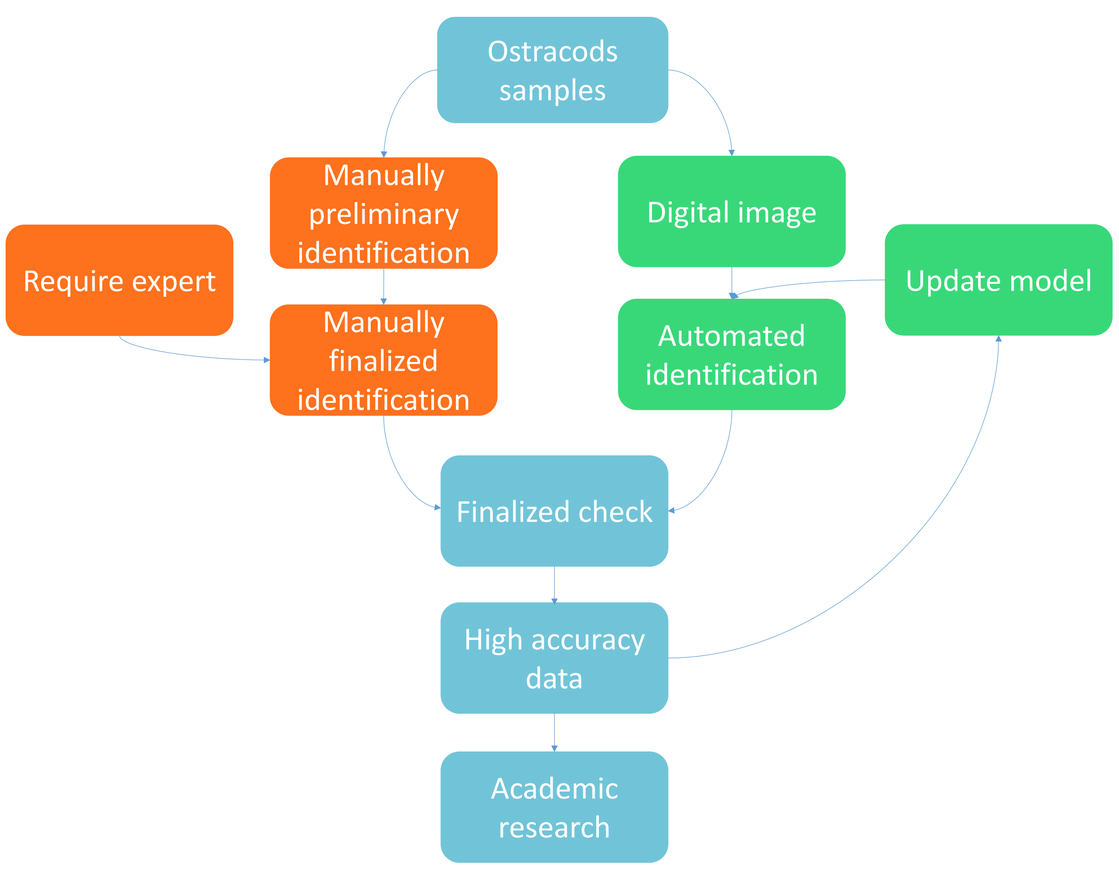
Automatic identification for paleoclimatology: Preliminary study

Jiamian Hu*, Yihua Chen*, Yuanyuan Hong* & Moriaki Yasuhara*
The University of Hong Kong

Background

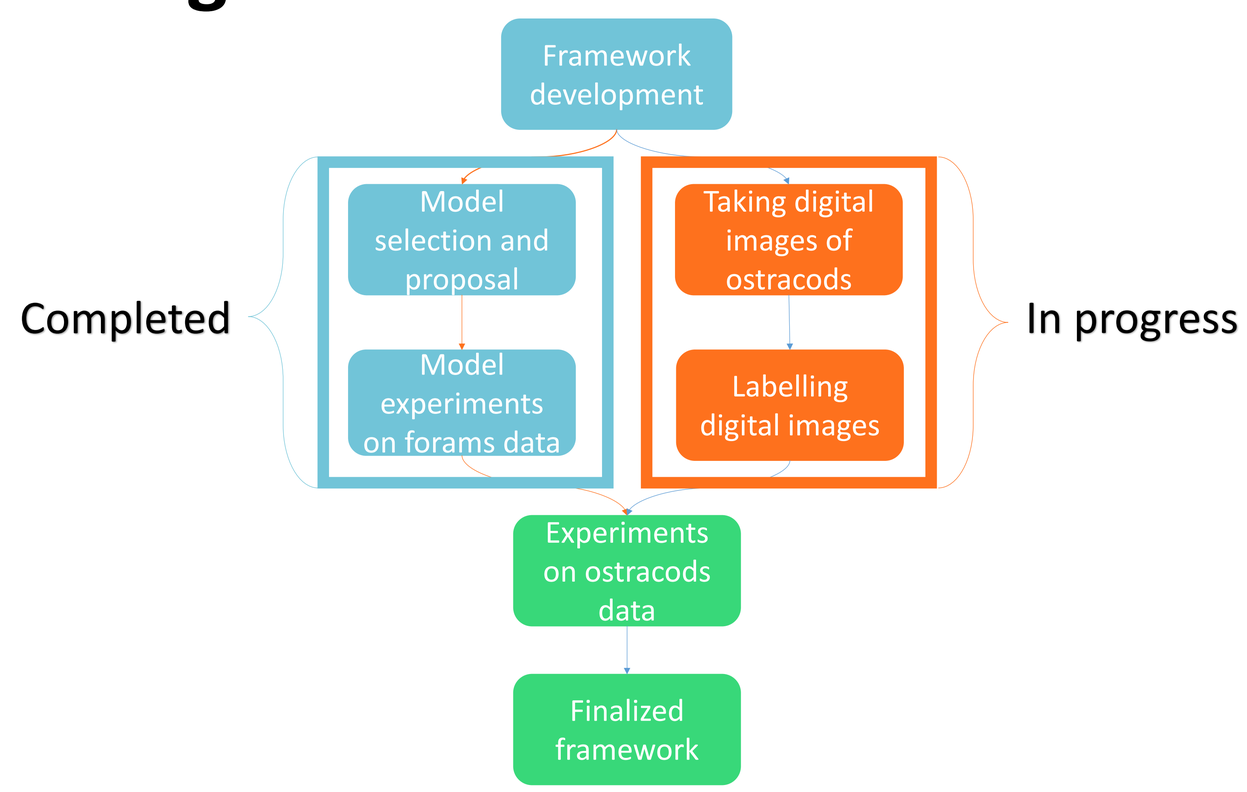
Ostracods are the microscopic remains of organisms and acted as an important proxy for paleoclimate reconstruction. However, identification of ostracods species, which is essential not only to know faunal composition but also chemical composition is difficult and requires substantial training and experience. Recent rapid development of deep-learning based image classification and object detection is providing the technology basis for automating the identification process using ostracods images acquired by the digital microscopes. In this project, we explored the automated taxonomy workflow by integrating the rapid developed deep-learning technology into the research pipeline. The test was conducted on Endlessforams dataset to show the effectives of new method.

Workflow



- Traditional workflow (orange color) require expert to do a finalized identification to confirm the identification result.
- Using automation workflow (green color) with deep learning models, we could train models identifying each species with reliable accuracy. New processed sample could be used to train updated models.
- Over 30,000 ostracods samples are prepared for training the models.

Progress



- We are taking digital images of ostracods samples using Keyence VHX-7000 microscope. 324 slides with each containing 90 ostracods on average will be recorded. Labelling of each ostracod is on-going.
- Model selection and test was completed using foraminifera image data with similar property of intended ostracods image from Endlessforams.

Experiments

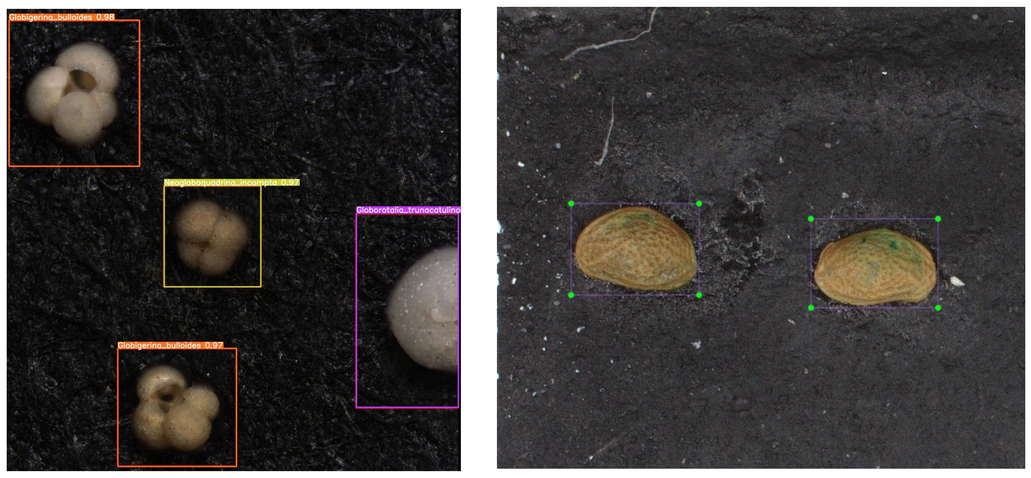
Experiments of 8 selected classification models are conducted on the Endlessforams dataset to test the reliability of adapting deep learning. In species classification task, the test accuracy of Rgenet-16-gf reached 88.8%. The average identification accuracy for each species in test set also reached 83.9%. For genus identification task, all test models achieved over 90% identification accuracy.

Classification	Species			Genus		
Model	**Acc.**	**AP**	**AR**	**Acc.**	**AP**	**AR**
ResNet-152	0.883	0.817	0.753	0.928	**0.928**	**0.832**
ResNet-50	0.870	0.838	0.740	0.916	0.881	0.754
VGG-16	0.866	0.774	0.753	0.922	0.831	0.776
VGG-19	0.873	0.825	0.764	**0.931**	0.871	0.821
Effieient-b4	0.844	0.817	0.646	-	-	-
Regnet-16-gf	**0.888**	**0.839**	0.756	-	-	-
ViT-b-16	0.836	0.825	0.686	-	-	-
Convnext-large	0.836	0.685	**0.825**	-	-	-

Detection	Species			
Model	**AP**	**AR**	**mAP@[.5:.95]**	**mAR@[.5:.95]**
Yolo-v5-l	0.871	0.860	0.891	0.890

As for object detection, Yolo-v5 achieved satisfied result by achieving both AP and AR over 80% on composited image of foraminifera.

Show Case



- Left image: Yolo detection results on raw slides of Endlessforams dataset.
- Right image: Labeled ostracods data taken from VHX-7000 microscope.

References

Hong, Y., M. Yasuhara, H. Iwatani, and M. Mamo. 2019. Baseline for ostracod-based northwestern Pacific and Indo-Pacific shallow-marine paleoenvironmental reconstructions: ecological modeling of species distributions. Biogeosciences 16:585–604.

Hsiang, A.Y., A. Brombacher, M.C. Rillo, M.J. Mleneck-Vautravers, S. Conn, S. Lordsmith, A. Jentzen, M.J. Henehan, B. Metcalfe, I. Fenton, and others. 2019. Endless Forams: >34,000 modern planktonic foraminiferal images for taxonomic training and automated species recognition using convolutional neural networks. *Paleoceanography and Paleoclimatology* 34:1157–1177.

Foram data：http://endlessforams.org/publications.html

PROJECT CODE: Automatic identification for paleoclimatology: Preliminary study