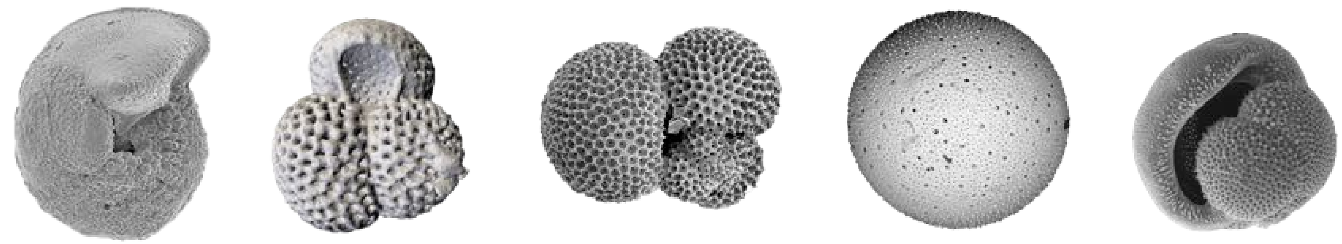
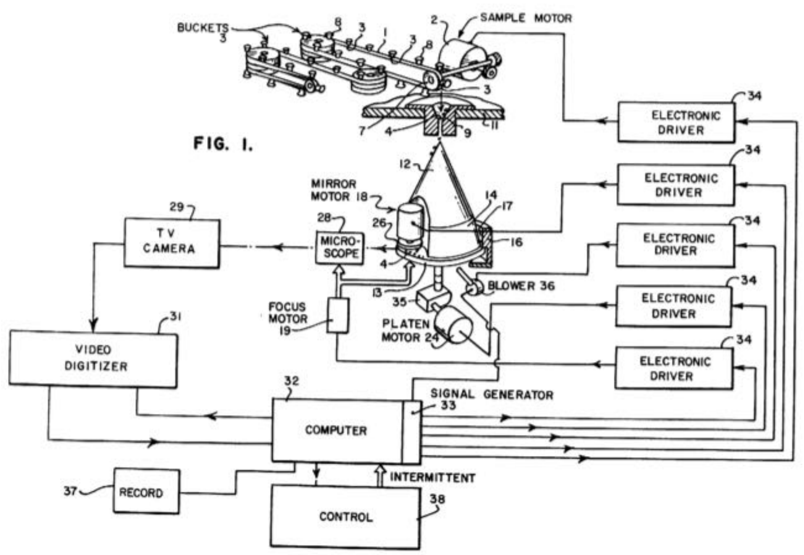
EGU 2020





Johnson, PhD UCSD, 1980

AUTOMATED RECOGNITION AND PICKING OF FORAMINIFERA USING THE MISO (MICROFOSSIL SORTER) PROTOTYPE

THIBAULT DE GARIDEL-THORON, ROSS MARCHANT*, MARTIN TETARD, MICHAEL ADEBAYO, AND YVES GALLY | CEREGE, AIX EN PROVENCE

@TdeGaridel

MISO - MICROFOSSIL SORTER

In collaboration with ATG Technologies, we designed and built a fully automated machine able to image and pick single particles in the 100 μm to 1mm size range.

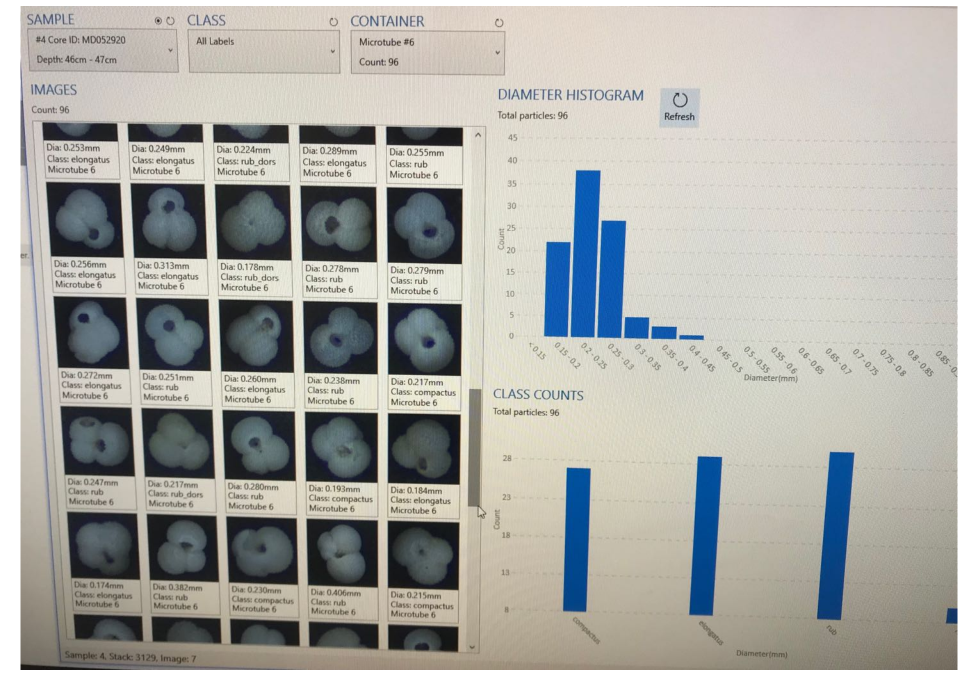
R. Marchant, post-doc, signal processing

ATG TECHNOLOGIES

MGA TECHNOLOGIES

https://www.atg-technologies.fr

Control software (MiSo)



Input → → Output

x 30

Sieved sample

~8000 particles/day

Z-stack imaging

CNN classification

x 120 microtubes

- Paleothermometer Mg/Ca, Δ47
- ^{14}C Datations
- Stable Isotopes
- Paleocirculation ε_{Nd}

x 3960 monospecimens

- Individual Foraminifera Analysis

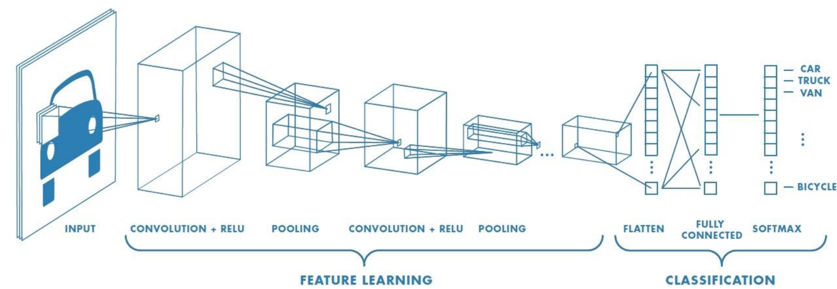
Patent (pending), 2 software

Particles are recognized using a Convolutional Neural Network classifier and separated in microtubes or micro slides. Machine is cleaned between each sample.

CONVOLUTIONAL NEURAL NETWORKS : A FORAM CNN

« In the area of computer vision (…) deep artificial neural networks have reached superhuman capabilities on a wide range of visual recognition problems »

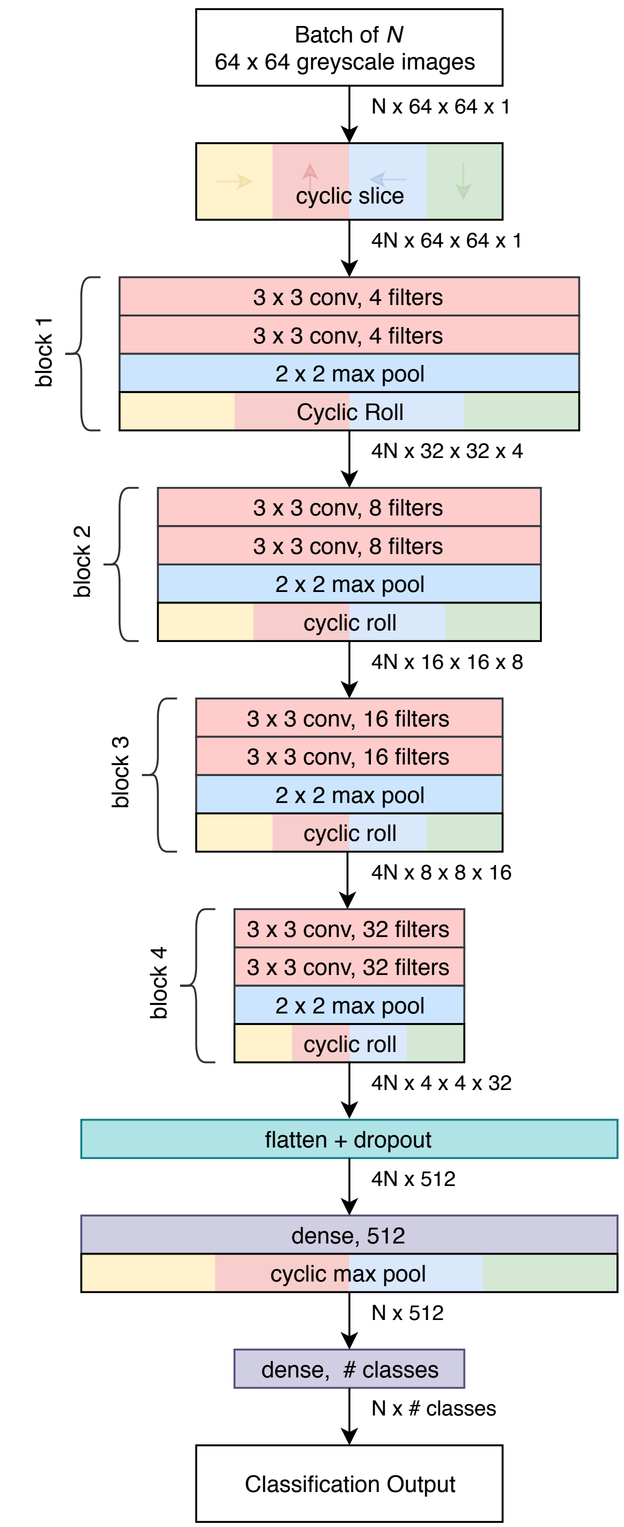
<u>Serre, Ann. Rev. Vision Sci., 2019</u>



MiSO

- CNN topology adjusts to input image dimensions
- Uses cyclic layers for rotation invariance
- Trained with augmentations common to microscope images

MiSo CNN



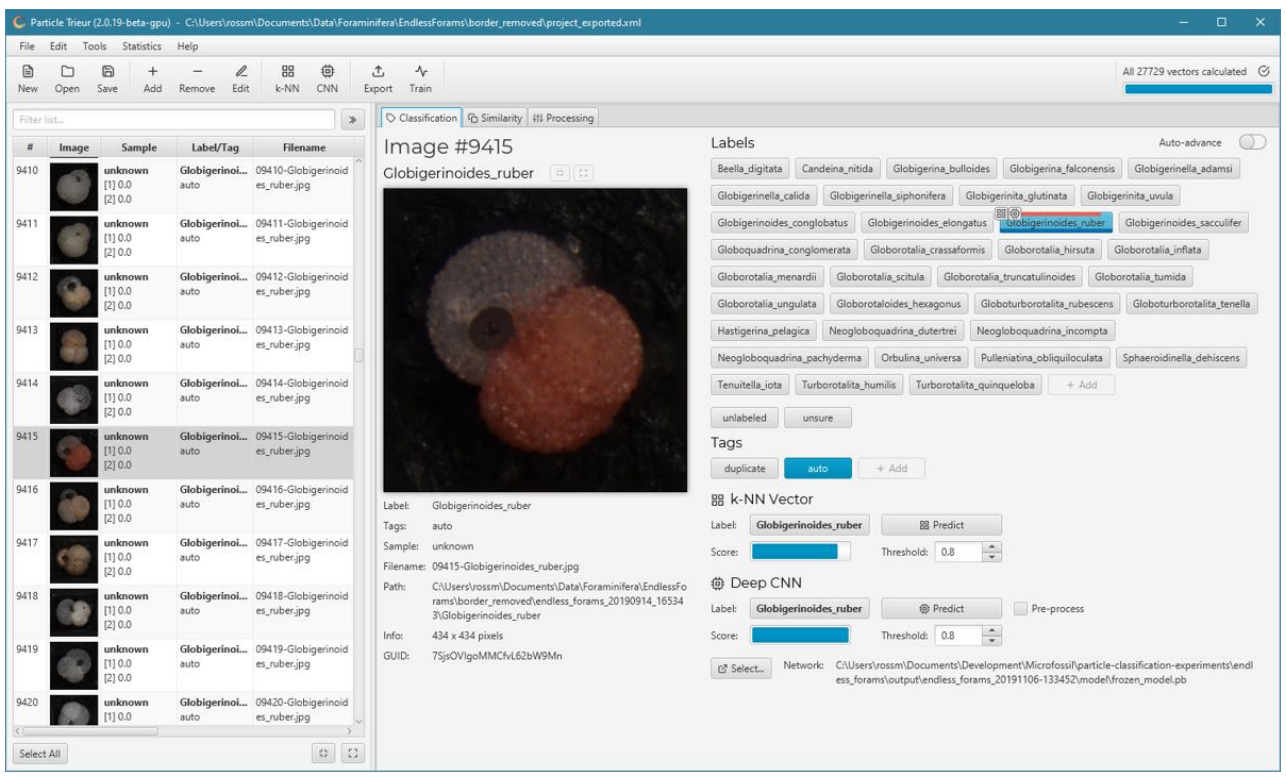
We tested a variety of CNN setups and are now using ResNet and BaseCyclic as in the preprint linked. The steps from the labeling to the training of the CNN on foraminifera images is achieved on a dedicated user friendly software ParticleTrieur.

<u>Read the preprint of Marchant et al., with the description of the tests for the CNN here</u>

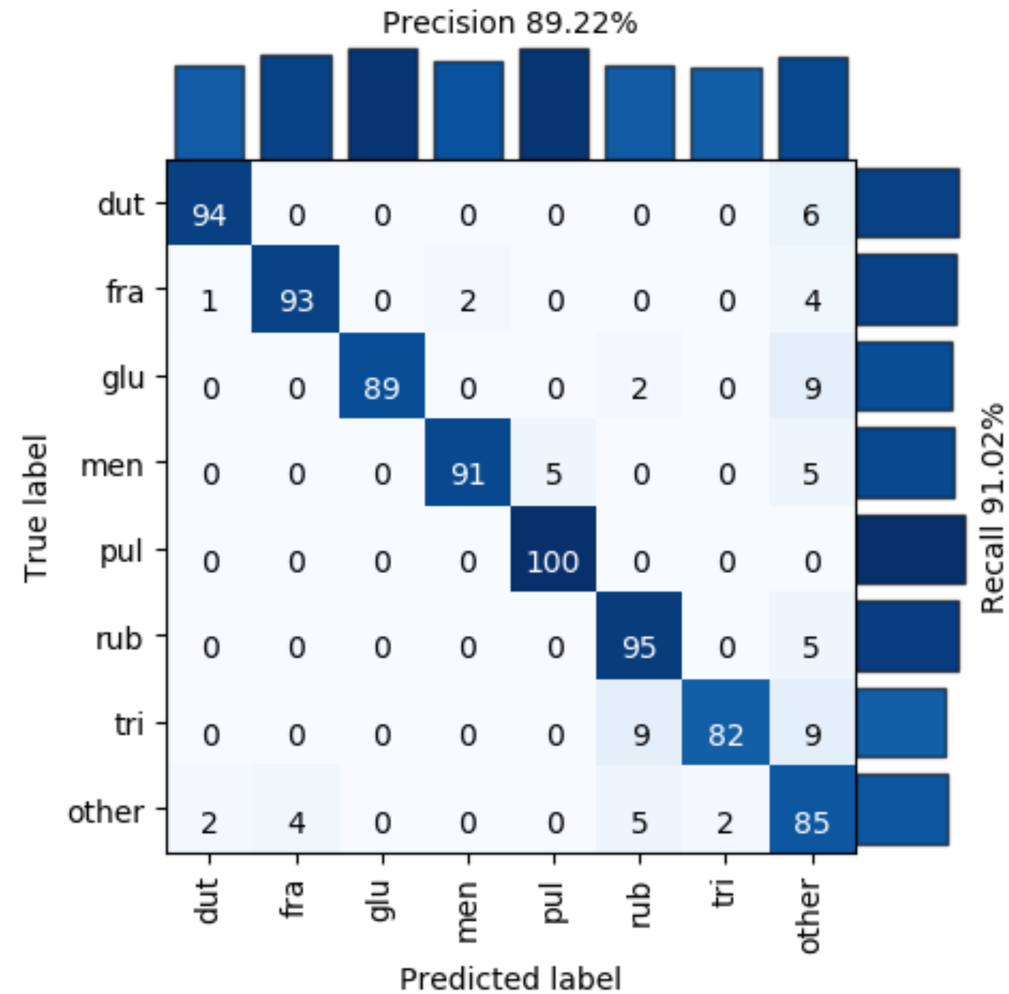
OUR WORKFLOW

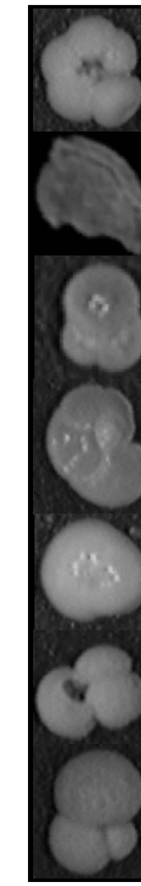
(1) Scan representative samples from the sediment core using MiSo (n>10,000 to 20,000 images)

(2) Label a training database using ParticleTrieur



(3) Train and evaluate a CNN classifier





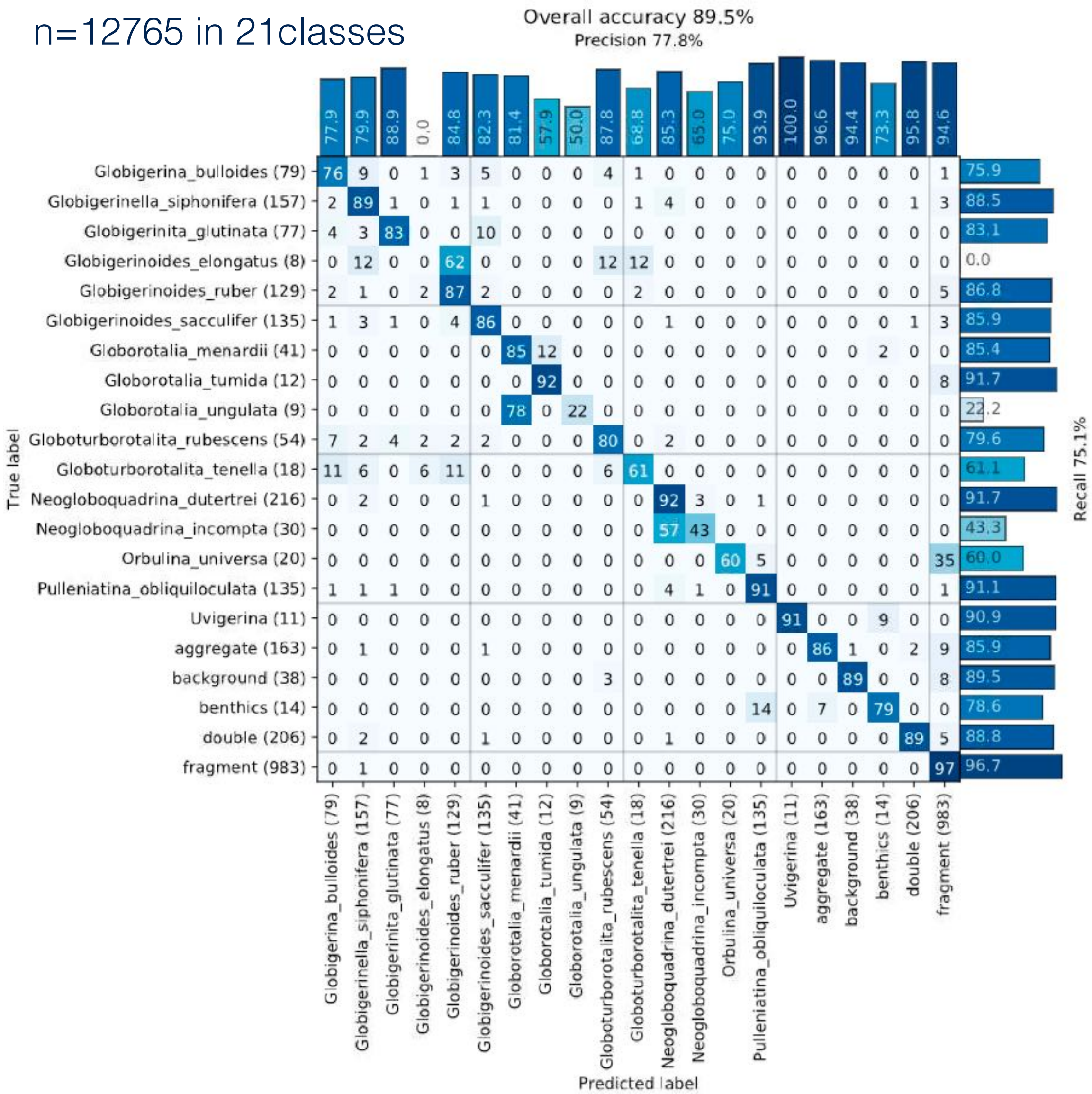
- Example of a simple training set (picking oriented) : each class includes different subclasses with different orientations
  - Typical precision & recall ~90±5%
  - Possible to cross-check later with different taxonomical frameworks
- Training set should be close to the working set
- Image pre-processing have to be very rigorously similar

For picking, when the number of false positives is more critical than false negatives, simple training as shown can learn to pick the most abundant classes with more >90% precision and recall

DOWNCORE RECONSTRUCTION

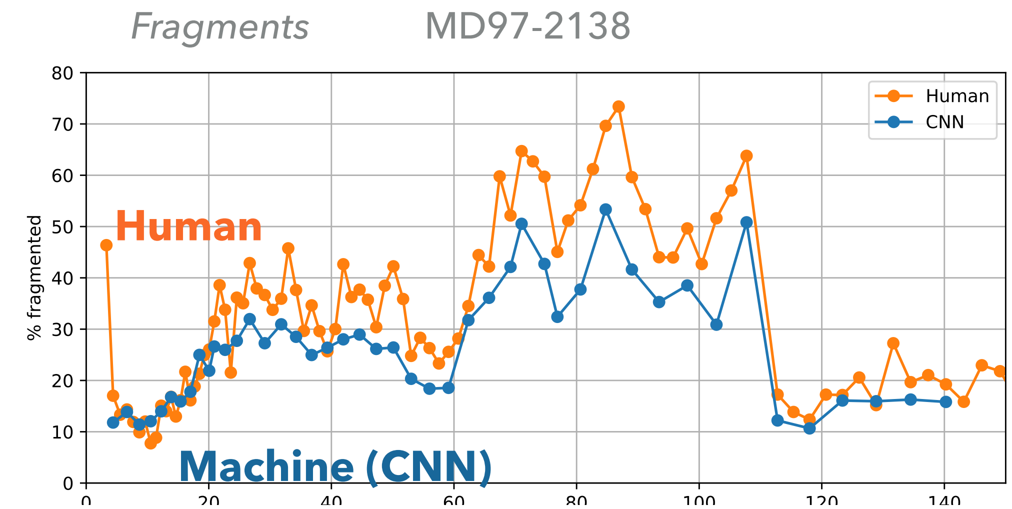
TRAINING

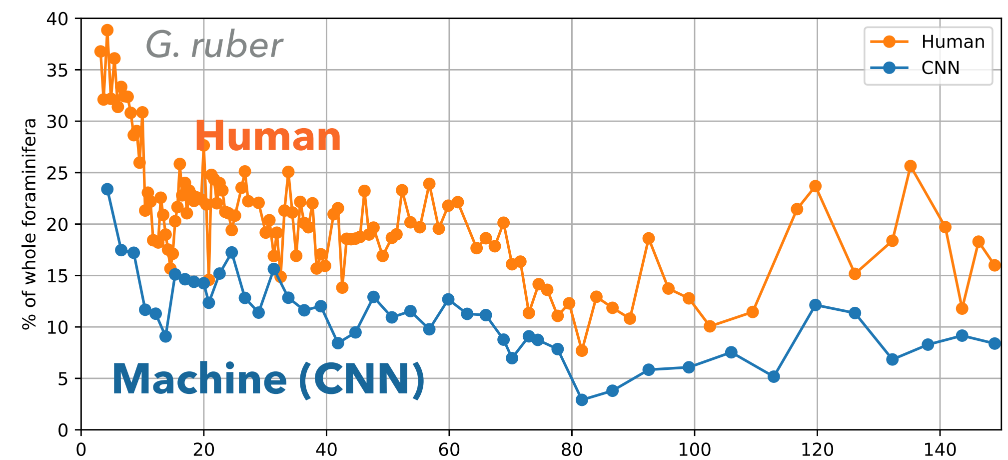
n=12765 in 21classes

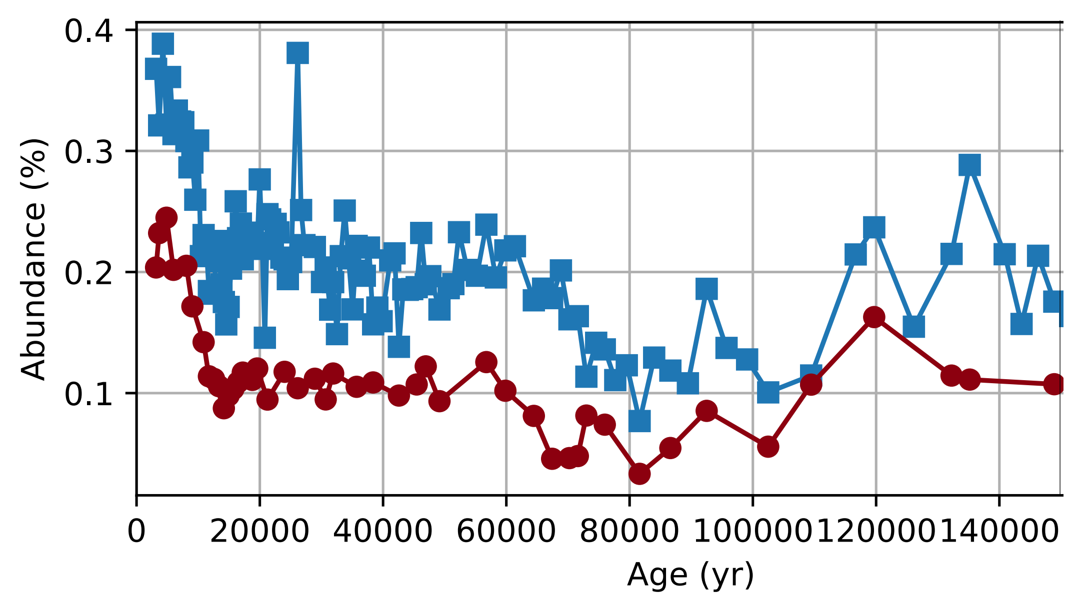


In the core MD97-2138, from the Western Pacific Warm Pool, we compare results from human counting with a CNN trained on this core. Fragment counts and the main species are close to the human counts. Yet, artefacts as clays infillings, dissolution artefacts can cause some misclassifications as in marine isotopic stage 5.

COMPARISON WITH HUMAN COUNTING







CONCLUSIONS

- Automatic imaging & recognition performed 24/7 routinely at CEREGE using CNNs
- Dedicated workflow for foraminifera from image acquisition to specimen handling
- Image preprocessing critical for image recognition
- Software packages (x-platform) user-friendly
- Ongoing developments : depth-reconstruction, size variations,
- Biometrics

- Very sensitive to acquisition method
  - Microscope lighting
  - Image resolution
- Data driven
  - Training set must span all variations
- Unknown forams can be incorrectly classified with high probability
- Size information is lost
- Misclassification causes signal offset and reduced range

Demo on request ! And if you're interested in crunching millions of images of sediments particles, get in touch.

What's next ? : PhD Michael Adebayo : Indian Ocean paleoceanography based on MiSo