Artificial intelligence approaches to optimising segmentation in computed tomography data processing

Wang YuHeng

December 2018

Supervisor: DR. MARTIN D. BRAZEAU

Affliation: Faculty of Natural Sciences, Department of Life Sciences (Silwood Park)

Email: m.brazeau@imperial.ac.uk

Supervisor: DR. MATTEO FUMAGALLI

Affliation: Faculty of Natural Sciences, Department of Life Sciences (Silwood Park)

Email: m.fumagalli@imperial.ac.uk

1 Key words

Computed Tomography; Segmentation; Fossil; Artificial Intelligence; Deep Learning; Convolutional Neural Networks

2 Introduction

Computed tomography(CT) is now widely used in palaeontology to study fossils. It can produce 3D virtual model of fossils from 2D x-ray images. It is a non-destructive method and needs little or no sample preparation(Dierick et al., 2007). With great improvement of CT technology, the bottleneck of CT-based analyses transfers from scanning to data processing(Abel et al., 2012). The first step of data processing is segmentation, which means partitioning of an image into non-overlapping, constituent regions that are homogeneous with respect to some characteristic such as intensity or texture(Pham et al., 2000). Traditional segmentation includes four steps: contrast enhancement; surface determination; region growing and masking tools(Abel et al., 2012). During the procedure of segmentation, manual selections (and therefore subjectivity and error) are inevitable and time-consuming because of the diminishing image contrast at the junction of fossil and matrix. As a consequence, following steps like rendering are easily influenced and tend to be challenging and error-prone. In this project, we prepare to use artificial intelligence(AI) based techniques to optimise the process of segmentation in fossil CT image data. Classifier and clustering are relatively mature machine learning techniques used in image segmentation, where Markov random field models(MRF) is often incorporated into them(Pham et al., 2000). Classifier is a supervised method which can be parametric or nonparametric while clustering is unsupervised. Currently, deep learning is the most popular machine learning method, where convolutional neural network(CNN) is most commonly applied in image processing. CNN is designed to process data in the form of multiple arrays like image data, containing convolutional layers and pooling layers as its unique structure(Lecun et al., 2015). It has advantages in processing label-abundant data including the segmentation of biological images(Ning et al., 2005). Due to the superiority of AI, especially CNN, in image processing, we prepare to use these techniques to replace manual selection. We expect it will address the problem of time consumption and error during segmentation to some extent.

3 Methods:

We will use 'off-the-shelf' AI frameworks like TensorFlow. We use an iterative approach. We will initially attempt to train AI algorithms to segment 'easy' datasets with very good contrast. These will include specimens of non-fossil material, or fossil material that has been chemically prepared and

has no surrounding rock matrix. We will then further improve on our methods by adapting them for fossils that have surrounding rock matrix, but nonetheless have good contrast. From here, we will make further refinements to try to adapt the methods to more problematic datasets where contrast is poorly normalised.

4   Anticipated outputs and outcomes

The results of this project will be compared with current "semi-automated" methods to see if it improves efficiency, quality or accuracy in process of segmentation.

5   Timetable

12/2018-02/2019	Studying deep learning and other basic knowledge; Reading literature about segmentation and convolutional neural net work; Writing the introduction of thesis.
02/2019-05/2019	Testing AI-based techniques in simple model; Replicating others algorithm to check the feasibility; Comparing results with traditional methods; Trying to optimise the algorithm and parameter.
05/2019-07/2019	Testing new algorithm in more complex model and optimise the parameter and structure of neural network; Comparing results with traditional methods; Writing the methods part of thesis.
07/2019-09/2019	Finishing the thesis about results and discussion; Polishing the thesis and correcting the mistake.

6   Budget

Lab equipment(CT scanning)	£200
High performance computing time	£150
Cost of support stuff(interpreters, translators, transcribers)	£150
Total	£500

References

R. L. Abel, C. R. Laurini, and M. Richter. A palaeobiologist's guide to 'virtual' micro-CT preparation. *Palaeontologica Electronica*, 15(2):496–500, 2012.

M. Dierick, V. Cnudde, B. Masschaele, J. Vlassenbroeck, L. Van Hoorebeke, and P. Jacobs. Micro-CT of fossils preserved in amber. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 580(1 SPEC. ISS.):641–643, 2007. ISSN 01689002. doi: 10.1016/j.nima.2007.05.030.

Y. Lecun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015. ISSN 14764687. doi: 10.1038/nature14539.

F. Ning, D. Delhomme, Y. Lecun, F. Piano, L. Bottou, P. E. Barbano, F. Ning, D. Delhomme, Y. Lecun, F. Piano, L. Bottou, and P. E. Barbano. Toward Automatic Phenotyping of Developing Embryos from Videos EDICS categories: 2-SEGM, 2-NEUR, 2-NFLT. *IEEE Transactions on Image Processing*, 14(9):1360–1371, 2005. URL `https://hal.archives-ouvertes.fr/hal-00114920/file/IEEE.pdf`.

D. L. Pham, C. Xu, and J. L. Prince. Current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, 2(1):315–337, 2000. doi: 10.1146/annurev.bioeng.2.1.315. URL `https://doi.org/10.1146/annurev.bioeng.2.1.315`. PMID: 11701515.

"I have seen and approved the proposal and the budget"

Name: DR. MARTIN D. BRAZEAU

Date: 09/12/2018

Signature: