Geochronology Discuss., referee comment RC2
https://doi.org/10.5194/gchron-2022-12-RC2, 2022

Comment on gchron-2022-12

Taryn Scharf (Referee)

Referee comment on "Technical Note: colab_zirc_dims: a Google-Colab-based Toolset for Automated and Semi-automated Measurement of Mineral Grains in LA-ICP-MS Images Using Deep Learning Models" by Michael C. Sitar and Ryan J. Leary, Geochronology Discuss., https://doi.org/10.5194/gchron-2022-12-RC2, 2022

The manuscript of Sitar and Leary is well written, includes thorough explanations of methods and results, with appropriate literary support. The construction of colab_zirc_dims demonstrates good working knowledge of those sectors of deep learning and computer vision that are applicable to the segmentation of reflected light images of zircon grains. This work thus provides the geological community with a valuable step forward in the development of highly accurate, rapid and automated tools for zircon image segmentation and shape measurement. I believe this manuscript should be published once comments have been addressed.

Specific Comments

- Line 45-46: Colab_zirc_dims works exclusively on reflected light (RL) images of zircons mounted in resin. The authors mention that zircon shape may be partially obscured by resin, resulting in minimum shape measurements instead of true dimensions. As colab_zirc_dims is presented as a tool for zircon shape measurement, could the authors kindly expand the discussion to cover whether the error introduced by reflected light images is significant, and whether or not it predisposes colab_zirc_dims to certain use cases? For example, do we know what proportion of a dataset is typically affected by this phenomenon? Is there any risk in comparative studies in which mounts have been differently handled (e.g. ground to different depths) or where shape measurements have been extracted from a variety of image types?
- Line 72-76: Please include an image that compares the segmentation achieved by traditional methods such as Otsu thresholding against those of colab_zirc_dims, when artefacts are present (e.g. anomalous bright spots, bubbles), to support this assertion. Alternatively, please include supporting literature references.
- Line 88: I am perhaps confused by the term “zonal area” – does this refer to the banding seen in cathodoluminescence (CL) images of zircon? Have the authors used AnalyZr to extract bands from within grains? Unfortunately, as AnalyZr was not developed for CL images, full grain segmentation from CL images is expected to fail. Perhaps reword the text to clarify, as it might mistakenly be interpreted as a recommendation to use AnalyZr for CL image grain segmentation.
- Potentially inconsistent terminology. Do the authors intend these terms to have different meanings, or do they all refer to a MaskRCNN implementation with FPN, using a ResNet backbone? If the latter, I’d recommend that terminology be standardised

throughout the paper.
Mask RCNN FPN (line 139 & 147)
Mask RCNN (line 151, Fig 2 caption)
Mask RCNN ResNet-FPN (line 178)

- Line 155: Table 1: The authors have selected training iterations of 4000-7000. Fig 4 shows that models stabilise at approximately 2000 iterations. Could the authors please include their reasoning for selecting model checkpoints at such high iterations? Is there any risk that these models are comparatively overtrained (meaningfully less generalised) than those around ~2000 iterations (e.g. how do they compare on the test dataset used for Table 3)?
- Line 184-185: The meaning of "sample-dependent…resolutions…(194 by 194 pixels…)" was not clear to me. Does this refer to the "max_zircon_size" criterion in the "mosaic_info" data table of the "Data Matching and Preparation" Colab notebook (lines 262-266)? Consider rewording the text to clarify.
- Lines 184-188: Could the authors please provide an indication of the nature of these zircon grains (e.g. sources, ages, sedimentary environment, histograms of shape parameter variation etc.) so that the reader has an understanding of how diverse the image test and validation datasets are? As the authors are using a small dataset to train deep convolutional neural networks, a reader may wonder how generalised the trained models are, and whether they will perform as well on zircon grains from different regions. Alternatively, if the authors feel that the small training dataset inhibits generalisation, please expand on this in the discussion.
- Line 190: Kindly indicate which of the images were hand-selected (perhaps rename the image files and refer the reader to the supplementary data).
- Lines 202-204: Please clarify what an iteration refers to, in this training regime. Additionally, consider specifying epochs and batch size in Table 1, for those readers who may wish to test the reported training strategy within their own Python implementation of MaskRCNN.
- Line 216: Kindly provide definitions, using simple terminology or mathematical formula, for the training mask loss and average precisions shown in Fig 4. Is there a reference for the source of these definitions that could be provided?
- Line 259, Section 4.3.1 Dataset Preparation tools: Colab_zirc_dims has a unique work flow with specific data and metadata requirements. I suggest including a flow diagram illustrating the segmentation and shape measurement procedure, inputs and outputs, including detail on image size and channels. This would help the reader understand the data requirements and process flow of colab_zirc_dims.
- Lines 277-280: Colab_zirc_dims is designed for data output by facilities using the Chromium software. It therefore has specific data and metadata requirements. There appears some flexibility around metadata, which the authors touch on in lines 277-280. However, after reading this text I felt I lacked a clear understanding of (1) whether or not any reflected light image dataset could be adapted for use in colab_zirc_dims and (2) how this may be done (e.g. automatically generating the necessary metadata files from user inputs via script). Please could the authors expand on the flexibility and limitations surrounding the application of this tool to reflected light datasets in general. This would help readers quickly identify whether this tool can be used on their datasets. The flow diagram suggested in the previous comment may help in this regard.
- Line 305, Table 3. Are the authors using the term "spot" as a synonym for "grain" in "Average segmentation time per spot"? Additionally, the footnote describes the metric as the time required for image segmentation. Does the reader interpret this metric as segmentation time per image, per grain in the image, or per analytical spot (potentially more than one spot per grain) in the image?
- Line 305, Table 3: Please provide definitions, using simple terminology or mathematical formula, for each of the tabulated metrics. Is there an appropriate literature reference for the definitions, which could be provided?
- Line 305, Table 3: Please could the authors add a description of the test dataset to help the reader understand how similar the test dataset is to the training and validation

datasets. This adds additional context to the performance evaluation results.

- Line 309: The authors refer the reader to Fig 4 and Table 3, in which models are differently named. Kindly standardise model names throughout the paper, thus facilitating quick comparison of models across tables and figures.
- Line 310: Consider amending “training loss” to “training mask loss”, to be consistent with Fig 4.
- Line 335: Please clarify the meaning of “skew slightly negative”.

Technical Corrections

- Line 68: “with via” amend to “via”.
- Line 269: “…allows to users to generate…” amend to “…allows users to generate”.
- Line 292: “…are can…” amend to “can”.
- Line 323: “Centermaks2” amend to “Centermask2”