

Automatically linking Dead Sea Scroll transcriptions to fragment images : Towards the letter level

Bronson Brown-Devost (U Göttingen), Nachum Dershowitz (Tel Aviv U), Daniel Stökl Ben Ezra (EPHE, PSL)

One of the outcomes of a joint venture of the Scripta Qumranica Electronica [1] and the eScriptorium [2] projects is the future inclusion in the SQE database of the positions of lines, words and letters on the IAA photos of the fragments with a minimal number of complete letters. We used eScriptorium cum kraken's customizable layout segmentation method to locate columns and lines on the photos. Users can define both the layout typology as well as the stack of convolutional and recurrent neural network layers used in the segmentation process via the VGSL language [3]. Our segmentation process was iterative between automatic inference, manual correction of the errors of the automatic segmentation system, retraining a better segmentation model, applying it to a larger group of fragments and less time consuming manual correction. After the final manual correction, we created a rough automatic transcription and then aligned this rough transcription with the text from the database similar to the procedure described here [4]. We subsequently applied text to image alignment to find the word and approximate letter locations for each line. We have experimented with several methods of refining those approximate positions, such as SIFT-flow, to obtain precise bounding polygons in the image of each transcribed letter. Once this has been accomplished, a searchable pdf can be created with an image layer and a text layer for each fragment.

[1] Brown-deVost, Bronson, "Scripta Qumranica Electronica (2016–2021)," [HeBAI 5 \(2017\): 307–315](#).

[2] Stokes, P.A., B. Kiessling, D. Stökl Ben Ezra, R. Tissot, and H. Gargem. [The eScriptorium VRE for Manuscript Cultures](#). *Ancient Manuscripts and Virtual Research Environments*, ed. Claire Clivaz and Garrick V. Allen. Special issue of *Classics@* 18 (2021).

[3] Kiessling, B. A Modular Region and Text Line Layout Analysis System. *17th International Conference on Frontiers in Handwriting Recognition (ICFHR) (2020)*

[4] Stökl Ben Ezra, D., Brown-DeVost, B., Dershowitz, N., Pechorin, A., Kiessling, B., Transcription Alignment for Highly Fragmentary Historical Manuscripts: The Dead Sea Scrolls. *ICFHR (2020)* 361-366.