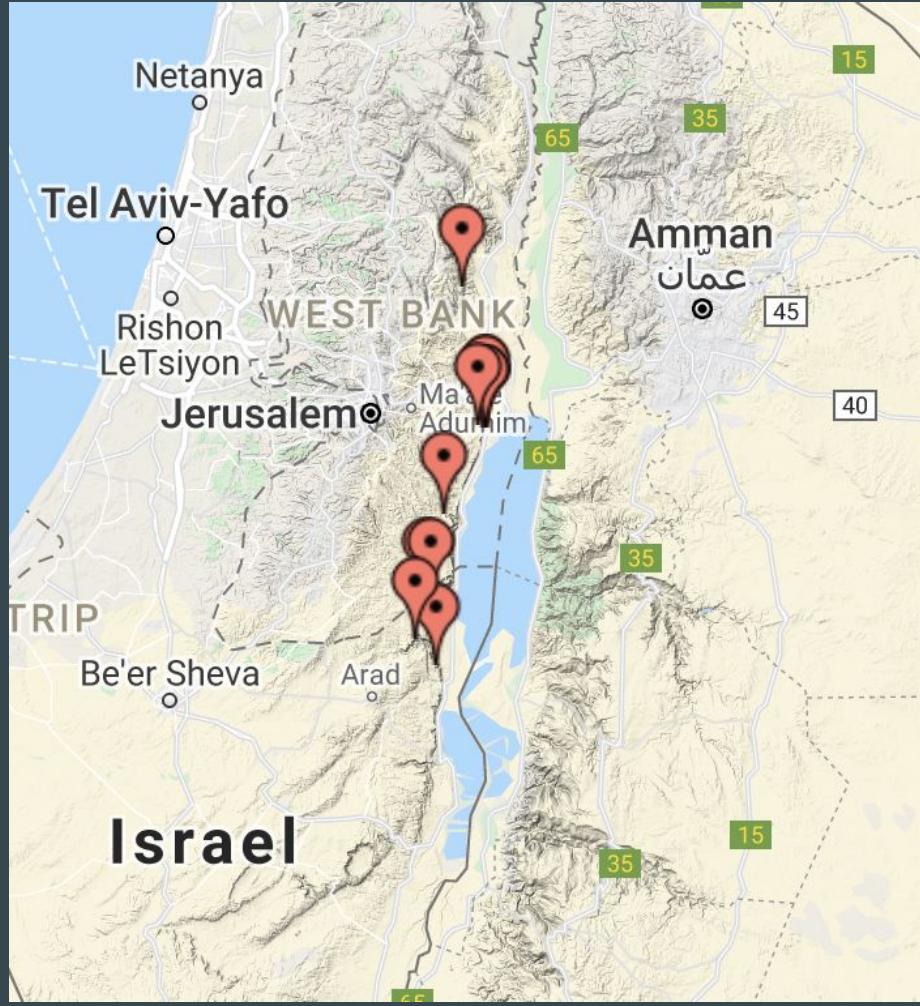


# The Dead Sea Scrolls

...

# Historical Relevance

- Before 1947 the oldest known biblical manuscripts were from the medieval period - but cave 1 changed that.
  - They are dated between third century BCE and the first century CE
  - unique vantage point for studying the latest literary evolutionary phases of what were to become the Hebrew Bible
- Till date they are the 2nd oldest
  - Oldest being : two silver scroll-shaped amulets containing portions of the Priestly Blessing from the Book of Numbers, excavated in Jerusalem and dated 600 BCE



## Qumran Caves

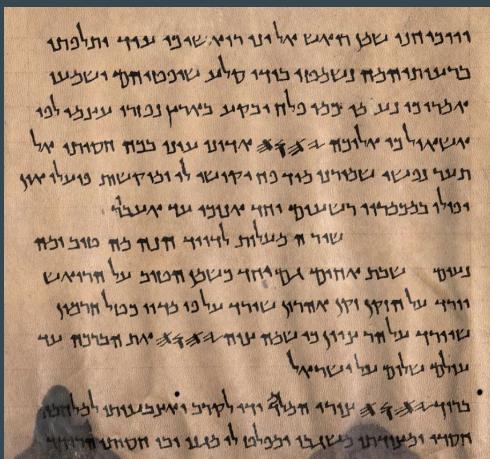
# What they contain

- One complete copy and 1 partial copy of the book of Isaiah
  - A retelling of the book of Genesis
  - Interpretation of the book of the old testament
    - Literary style not seen before
  - Rules and practises of a the religious community living there
  - Since writings from ancient israel were on papyrus it was a shock that they survived in the first place for that long
    - Cave 4 (1954) : 15,000 fragments found
    - International team to assemble them
      - Thickness, writing style etc
      - Used tape to stick it together
    - Most of the texts are written on parchment, some on papyrus, and one on copper.
-



- Story of one of the initial translators

# The 10 commandments



## The book of psalms

## The book of Genesis

In the beginning god created the heaven and  
the earth...



# The Church

- Initially they didn't give much importance to them
  - However soon people began to give comparisons between the "teacher of righteousness" found in Cave 1 and Jesus
    - An interpretation that the Teacher of righteousness had been crucified (now discredited)
    - Other interpretations that could pose as a challenge
  - Congregation for the doctrine of the faith(inquisition) got involved in it. To make sure nothing clashed with catholic doctrine !
-

# The Monk System

- University of Groningen - Research group ALICE Prof. Lambert Schomaker.
    - Collaborate with Qumran Institute, Faculty of Theology and Religious Studies, University of Groningen
  - The *Monk* system
    - HWR of historical documents
    - Word Spotting (google for HWR)
    - Dating of manuscripts
    - Geographic localization
    - Writer identification
  - 383 documents, 70k page scans
    - Qumran scrolls, Ming Qing Chinese poetry, Medieval charters 1300-1550 in 25yr periods, Russian handwritten newspaper 1672
-

1905 <sup>9<sup>o</sup></sup>  
Kantongerechtshof.

Maand 24/3/9 Rappt P. te Haarlem h. 54.4. kolt, te beschouwing  
het voorstel by het Kantongerecht te Haarlem  
tot W. H. A. Horsten.  
— Rechts geschied.

Maand 31/3/9 Rappt P. te Haarlem h. 5.15. om aan P. de Groot  
welke te Stadhuisen op zijn voorstel en Haar-  
lem te oordelen, integreer betrekking van Haar-  
lemmerse plaatvoornemers in het kantoor  
Bestuursraad en ziche niet daar betrekking  
Bestuursraad.

April 1/9 Rappt P. te Haarlem h. 5.4. kolt, te beschouwing  
het voorstel van P. te Haarlem h. 5.15. / te beschouwing Bladg 9  
tot W. H. Horsten.  
P. de Grootman tot te Haarlemmerse  
plaatvoornemers in het kantoor  
Horsten. W. H. S. Rapport. — Rechts geschied.

April 6/9 Rappt P. te April 6/9 kolt, te beschouwing  
het voorstel van de Haarlemmer plaatvoornemers  
te Haarlem h. 5.15. / te beschouwing h. 5.4.  
tot W. H. Horsten. Rapport. — Rechts geschied.

April 7/9 Rappt P. te April 7/9 kolt, te beschouwing  
het voorstel van de Haarlemmer plaatvoornemers  
te Haarlem h. 5.15. / te beschouwing h. 5.4.  
tot W. H. Horsten. Rapport. — Rechts geschied.

1905 <sup>9<sup>o</sup></sup>  
Kantongerechtshof.

April 9/9 Rappt P. te Haarlem h. 5.4. kolt, te beschouwing  
het voorstel van de Haarlemmer plaatvoornemers  
in het kantoor Haarlemmer, tot W. H. Horsten.  
Rapport. — Rechts geschied.

April 11/9 Rappt P. te April 11/9 om aan P. de Groot  
welke te Stadhuisen op zijn voorstel en Haar-  
lem te oordelen, integreer betrekking van Haar-  
lemmerse plaatvoornemers in het kantoor  
Bestuursraad en ziche niet daar betrekking  
Bestuursraad.

April 14/9 Rappt P. te April 14/9 om aan P. de Groot  
welke te Stadhuisen op zijn voorstel en Haar-  
lem te oordelen. — Rechts geschied.

April 14/9 Rappt P. te April 14/9 om aan P. de Groot  
welke te Stadhuisen op zijn voorstel en Haar-  
lem te oordelen. Rapport. — Rechts geschied.

April 16/9 Rappt P. te April 16/9 kolt, te beschouwing  
het voorstel van de Haarlemmer plaatvoornemers  
te Haarlem h. 5.15. / te beschouwing h. 5.4.  
tot W. H. Horsten. Rapport. — Rechts geschied.

April 21/9 Rappt P. te April 21/9 kolt, te beschouwing  
het voorstel van de Haarlemmer plaatvoornemers  
te Haarlem h. 5.15. / te beschouwing h. 5.4.  
tot W. H. Horsten. Rapport. — Rechts geschied.

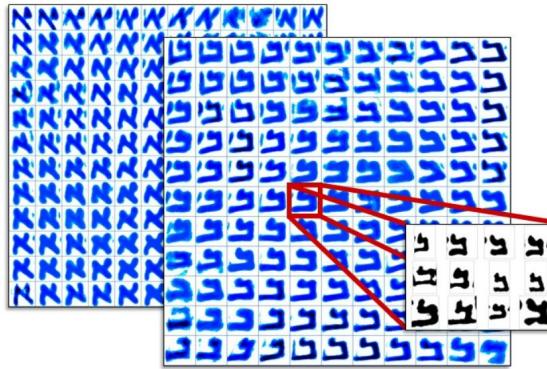


# Scribe Identification

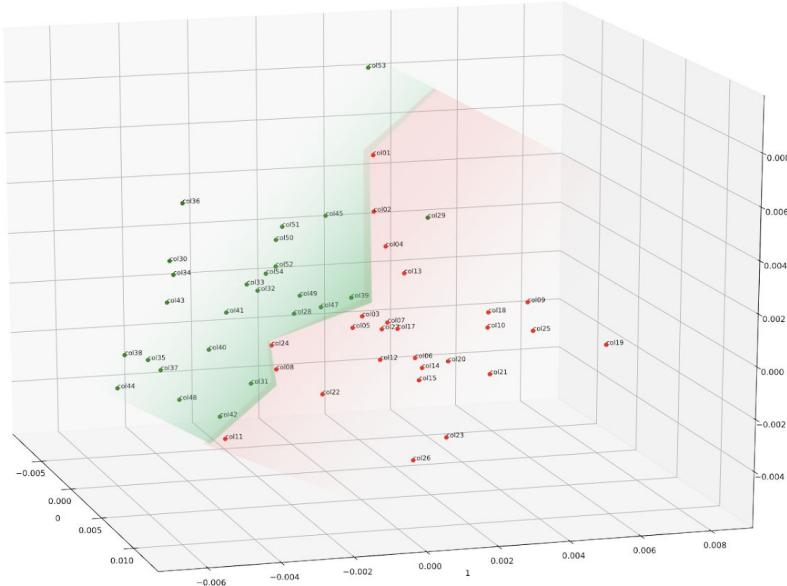
[Popović, M., Dhali, M. A., & Schomaker, L. (2020). Artificial intelligence based writer identification generates new evidence for the unknown scribes of the Dead Sea Scrolls exemplified by the Great Isaiah Scroll (1QIsaa). *arXiv preprint arXiv:2010.14476.*]

- One of the problems in traditional palaeography is to determine writer identity when the writing style is near uniform
  - Exemplified in the great Isaiah Scroll
  - Many scholars believe that it was written by one scribe
  - Coming up with an objective measure/evidence
- Concluded that 2 scribes collaborated on the manuscript and found a point of transition at column 27

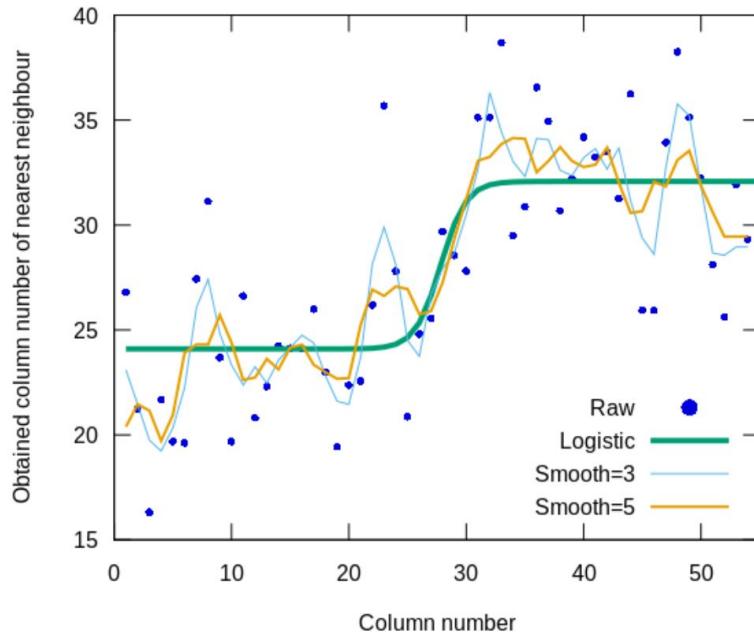




**Fig 1.** Two 12x12 Kohonen maps (blue colormaps) of (full) character *aleph* and *bet* from the DSS collection. Each of the characters in the Kohonen maps are formed from multiple instances of similar characters (shown with zoomed box with red lines). These maps are useful for chronological style development analysis. In our current study of writer identification, fraglets will be used instead of full character shapes to achieve more precise (robust) results.



**Fig 5.** Hinge feature plot (using PCA) of the full column images of 1QIIsa<sup>a</sup> with a clear line of separation (red: columns 1–27; green: column 28–54).

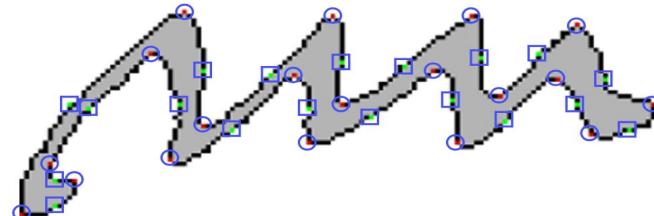


**Fig 9.** Average serial column position of the nearest-neighbour of a column, measured in fraglet-feature space. Raw samples (points), curves smoothed over 3 and over 5 points, and a best-fit logistic curve

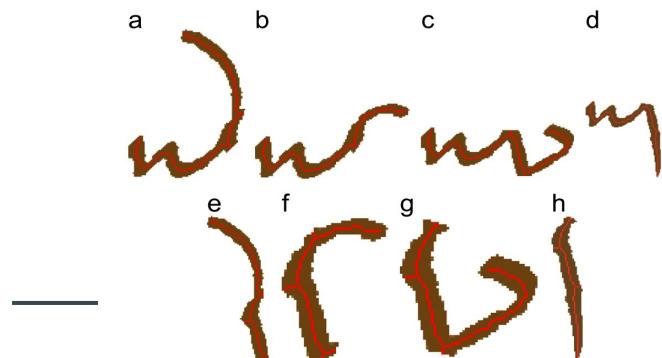
# Manuscript Dating Using Contour and Stroke Fragments

He, Sheng, et al. "Historical Manuscript Dating Using Contour and Stroke Fragments."

- A family of local contour fragments(kCF) and stoke fragment features(kSF) are discussed which can be used in dating of Historical manuscripts

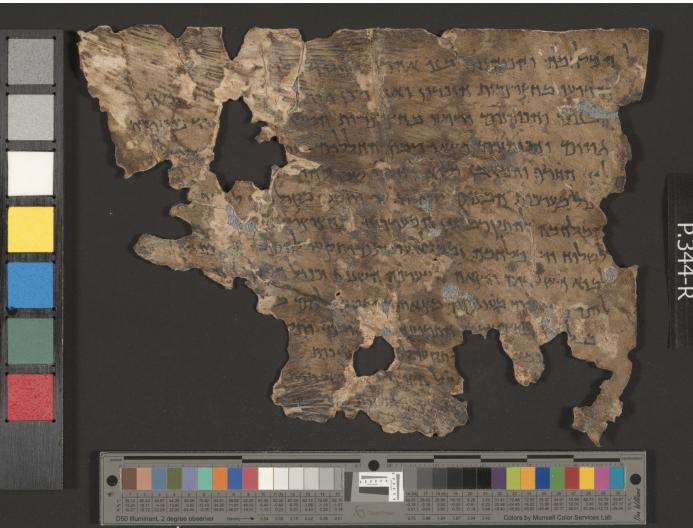


**Fig. 3.** A contour extracted on the connected component. The red points (with circle) are key points detected by the DCE [18] method and the green points (with rectangle) are the break points, necessary for capturing curvature information. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)



**Fig. 8.** Stroke fragments of 3SF generated in the SFG in Fig. 7(b). The corresponding nodes are: (a) {1,2,4}, (b) {1,3,4}, (c) {1,5,6}, (d) {1,5,6}, (e) {2,4,5}, (f) {3,4,5}, (g) {4,5,6}, and (h) {4,5,7}.

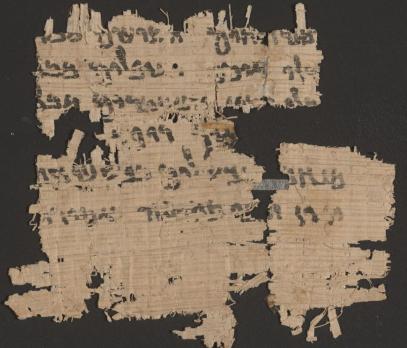
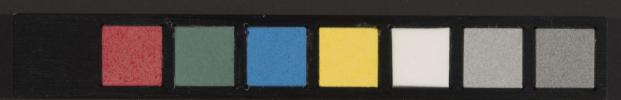
# What i worked on



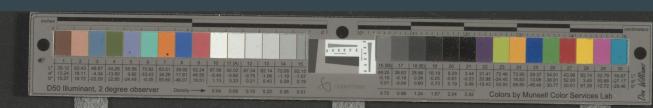
Text



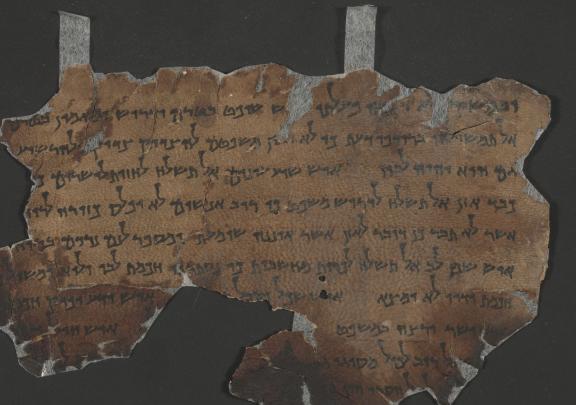
- A small dataset of example characters (cutout by the linguists who could clearly identify some of the letters)
- Some incomplete bigram data



P.22-R



P.106-R



P.123-R



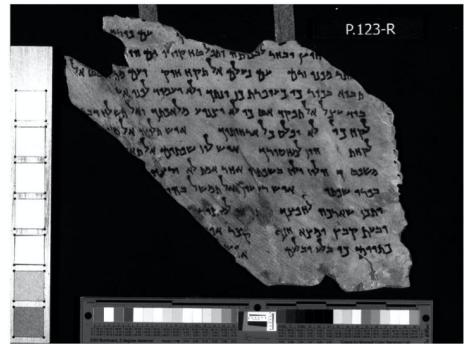
# Pipeline

- Word Segmentation
  - Parchment extraction : binarization, erosion, dilation, image reconstruction
  - Line segmentation : binarization, connected components to remove holes, histogram based segmentation
  - Word segmentation : ascenders, descenders, edge words, connected components, broken characters
- Character recognition
  - Data augmentation
  - Sliding window, LSTM based RNN

Some examples of the 1st part of the pipeline !

---

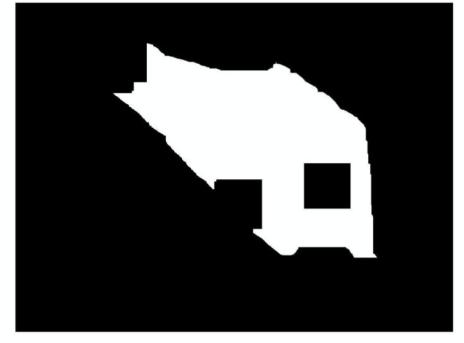
**Original Image**



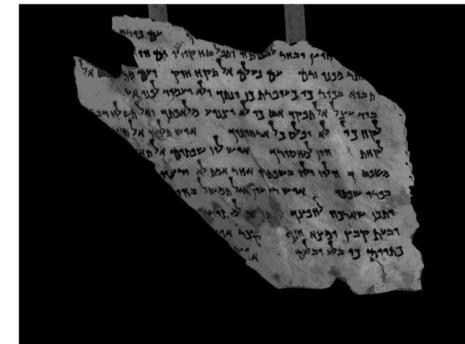
**Binarized Image - OTSU**



**Morphological Operations**

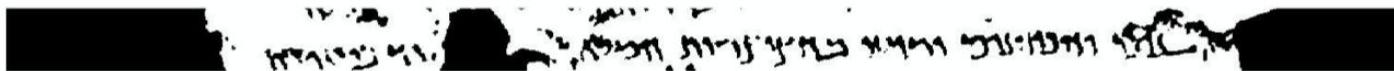


**Image Reconstruction**

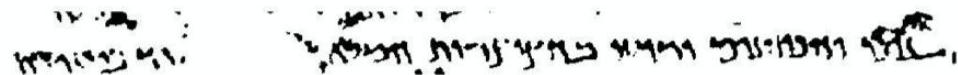




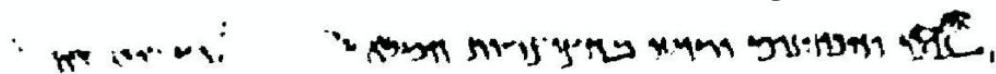
### **Segmented Binarized Line**



### **Remove Background Components**



### **Remove CC near the Boundary**



### **Segmented Words**



Thank You

Questions ?