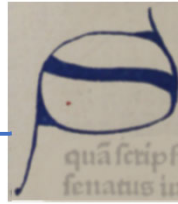


# Historia Scholastica at Boise State University



- Written by Peter Comestor in the early 1170s
  - A secular account of the Bible as a history, correlated with non-Christian events
  - A 15<sup>th</sup> century copy owned by Boise State University
  - Latin with many scribal abbreviations.
  - Hain's Repertorium Bibliographicum lists eleven editions, ten in Latin and one in French.
- 
- Project Goals:
  - Transcribe text
  - Translate text to English
  - Determine printer
  - Compare text across editions
  - Determine foundational (source) edition
  - Create virtual display to share results with a wider audience.

The Historia Scholastica Project (HSP) since 2011 has manually transcribed and translated 70 of the 608 pages from Latin to English. Tools from the field of document analysis will accelerate the process. For instance, Character spotting [Majid, IJDAR/ICFHR 2022] is not sensitive to touching characters, does not require the presence of blank rows between lines of text, and allows for the inclusion of special symbols not common in most recognition software. Illuminated capitals and section markers will be recognized separately in a similar process. Natural Language Processing techniques will be used to expand the scribal abbreviations. The translation from Latin to English will follow, which will be used for deeper analysis of the text to explore the impact it has had on society. We hope to stimulate discussion of other applications of document analysis to other corpora.

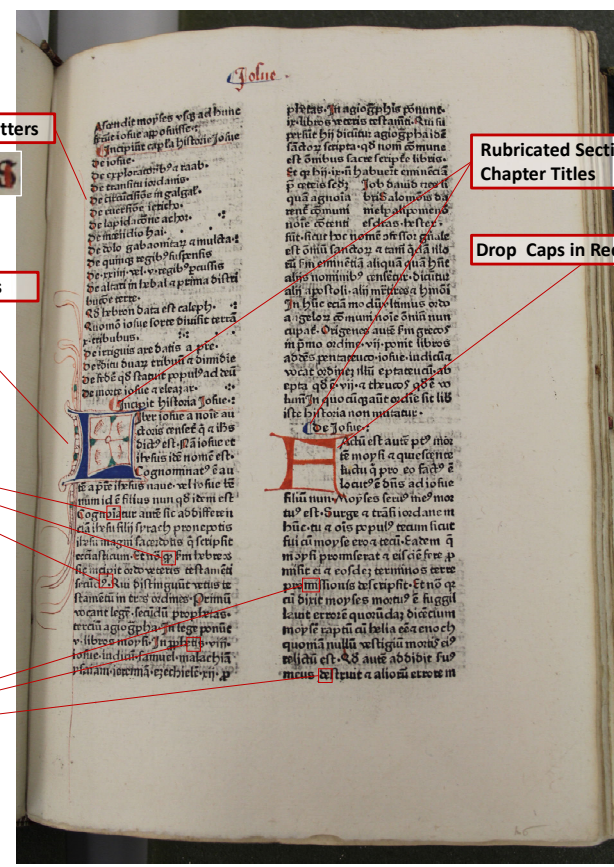
Rubricated capital letters



Rubricated Section & Chapter Titles

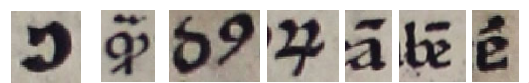
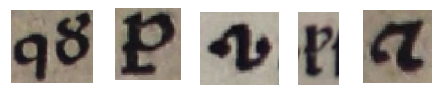
Drop Caps in Red & Blue

Illuminated capitals

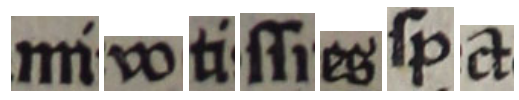


Scribal Abbreviations

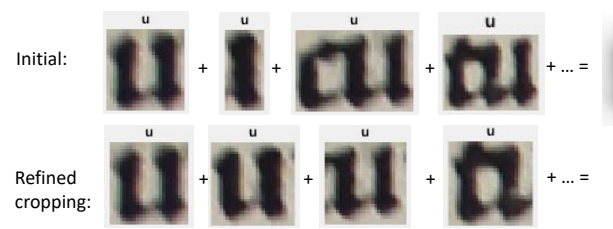
quod, per, id est, el, et, con, quam, dus, rum, an/am, bem, ere



Touching characters & ligatures



Character Prototypes for Training



Average shows improvement:

Recognition

- Issues:
- There are MANY symbols & abbreviations.
  - There are MANY ligatures.
  - Creating the training set will be very complicated. The provided transcript does not include the scribal abbreviations.

Image Processing



(1) Color separation  
 (2) Column segmentation  
 (3) Line segmentation  
 (4) Perspective Distortion correction

Additional Tasks:

- Page warp correction
- Drop Cap segmentation
- Capital Letter extraction
- Character shape template generation

Preliminary Results:

- Select 6 characters manually by selecting connected component.
- Form prototype character
- Refine bounding boxes
- Train YOLO object detector

