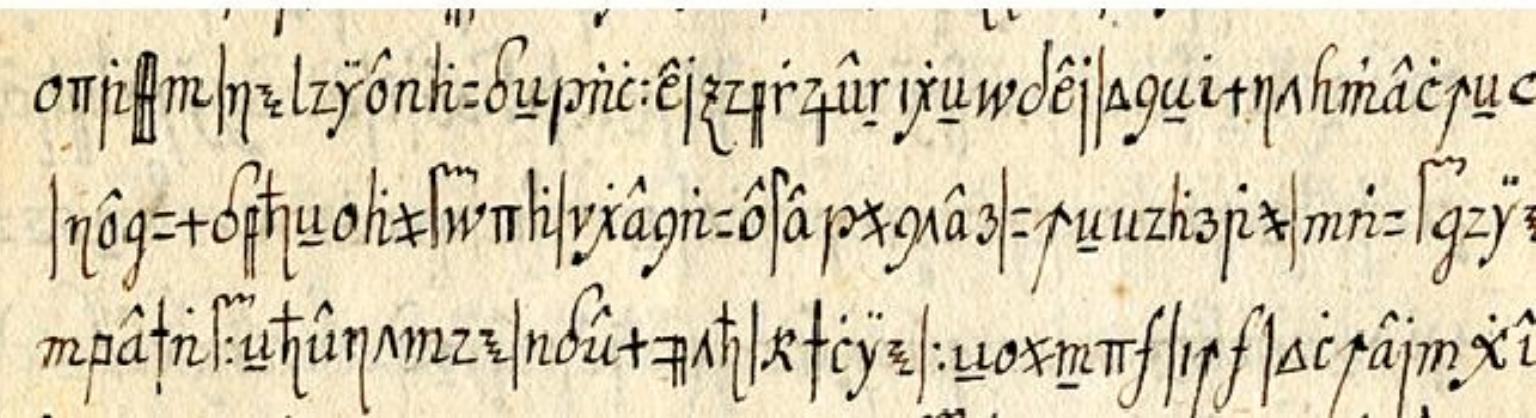
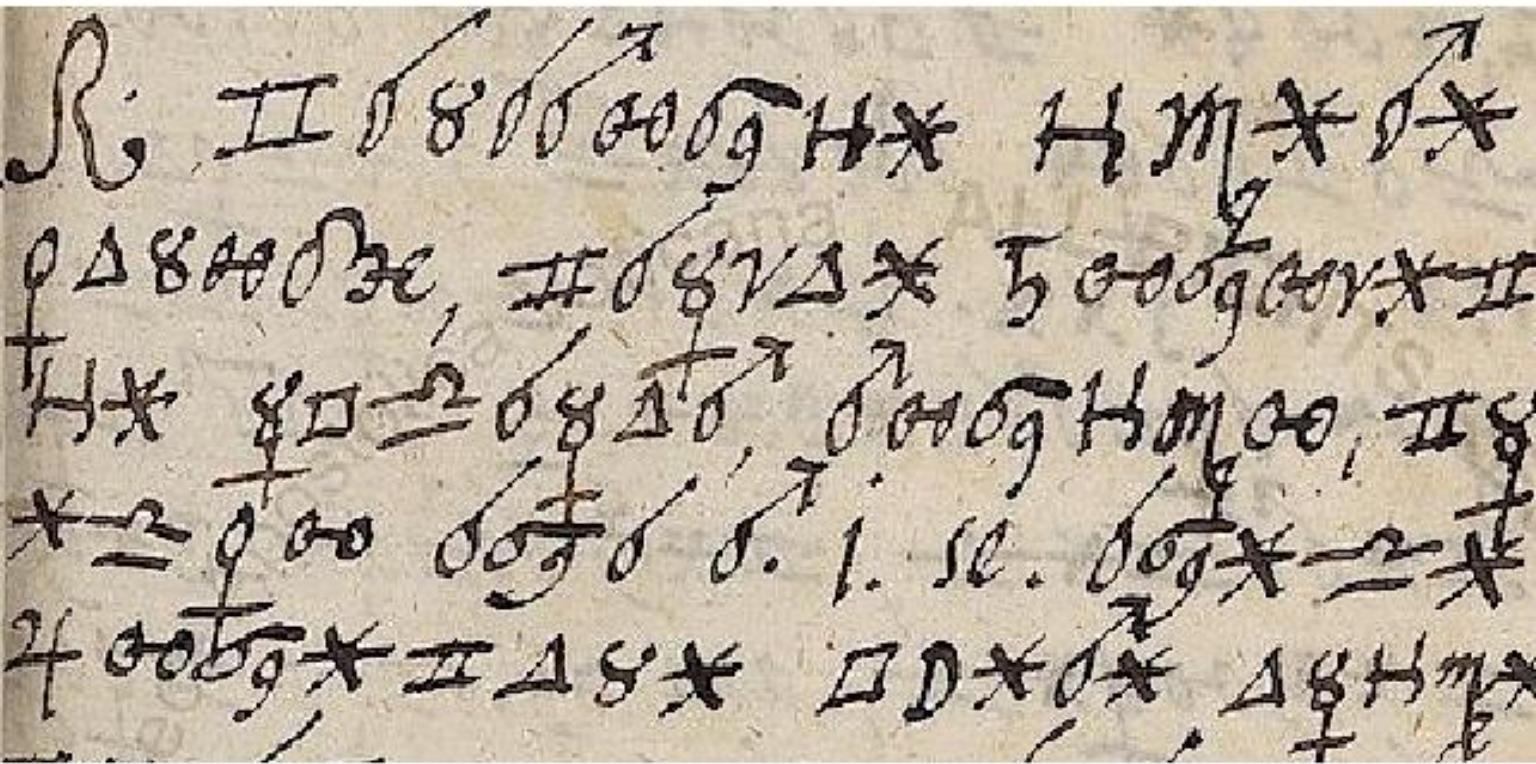
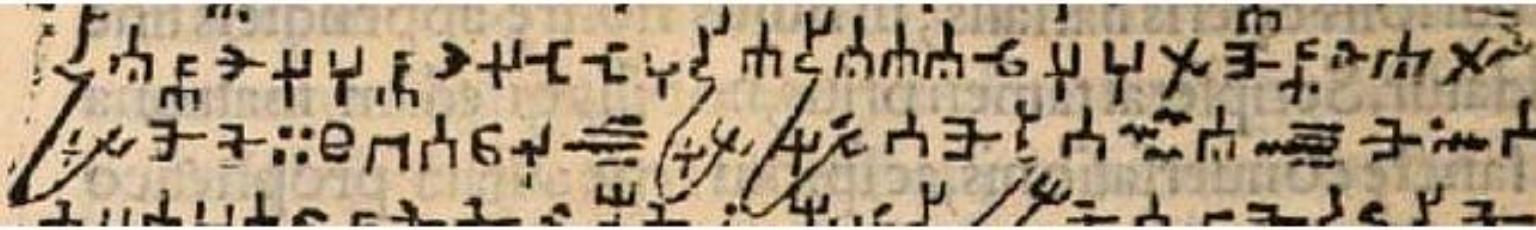


Handwritten Text (Character) Recognition

30.04.2024

Maria Levchenko / DHDK



References

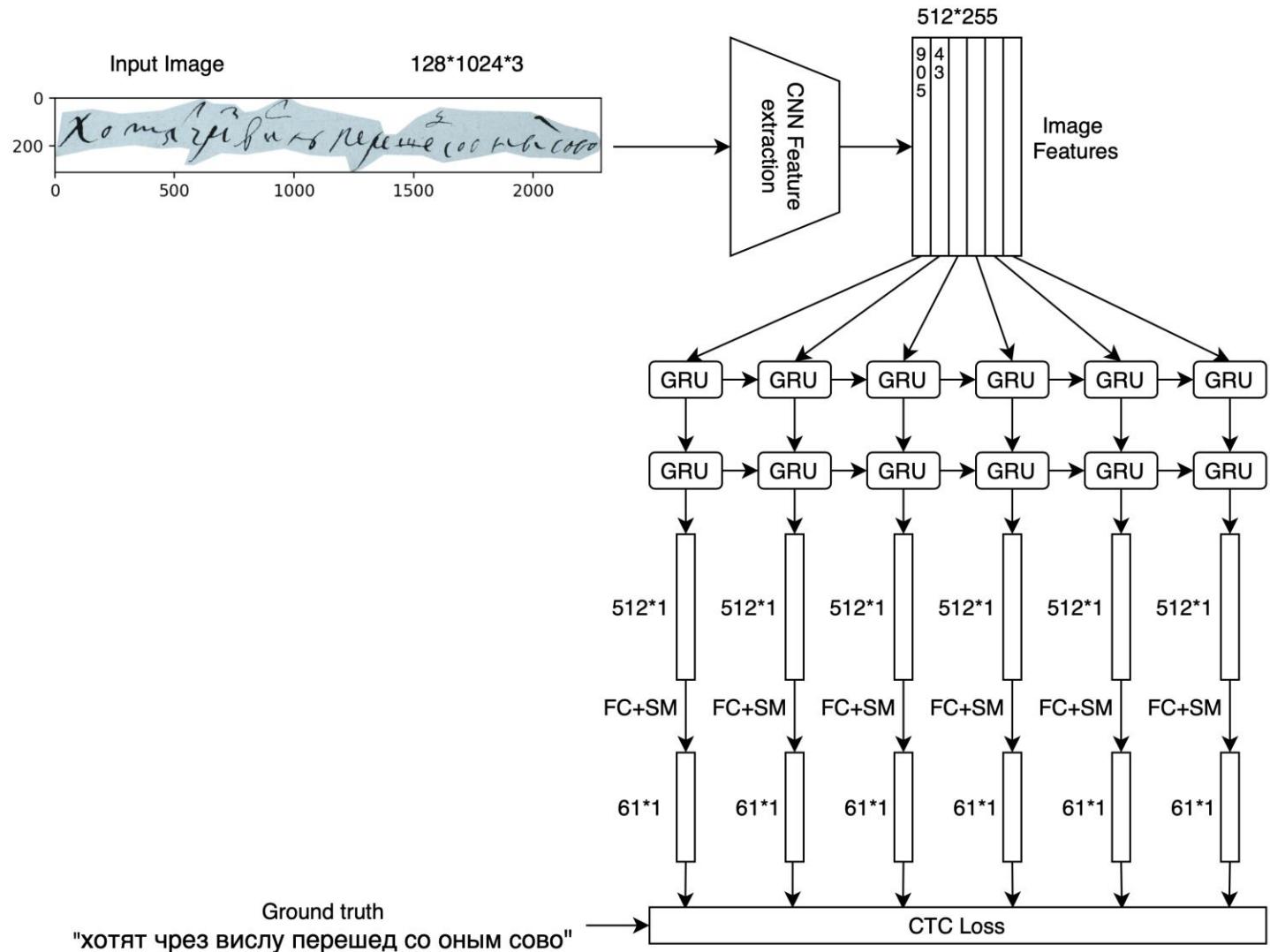
- Ströbel, Phillip Benjamin, Simon Clematide, Martin Volk, and Tobias Hödel. "Transformer-based htr for historical documents." arXiv preprint arXiv:2203.11008 (2022).
 - Muehlberger, G., Seaward, L., et all (2019), "Transforming scholarship in the archives through handwritten text recognition: Transkribus as a case study", Journal of Documentation, Vol. 75 No. 5, pp. 954-976. <https://doi.org/10.1108/JD-07-2018-0114>
 - Li, Minghao, Tengchao Lv, Jingye Chen, Lei Cui, Yijuan Lu, Dinei Florencio, Cha Zhang, Zhoujun Li, and Furu Wei. 2023. "TrOCR: Transformer-Based Optical Character Recognition With Pre-Trained Models". Proceedings of the AAAI Conference on Artificial Intelligence 37 (11):13094-102. <https://doi.org/10.1609/aaai.v37i11.26538>.
 - Najem-Meyer, Sven and Matteo Romanello. "Page Layout Analysis of Text-heavy Historical Documents: a Comparison of Textual and Visual Approaches." Workshop on Computational Humanities Research (2022).
 - Matteo Romanello, Sven Najem-Meyer, and Bruce Robertson. 2021. Optical Character Recognition of 19th Century Classical Commentaries:the Current State of Affairs.In Proceedings of the 6th International Workshop on Historical Document Imaging and Processing (HIP '21). Association for Computing Machinery, New York, NY, USA, 1–6. <https://doi.org/10.1145/3476887.3476911>
- Ayush Purohit et al, A Literature Survey on Handwritten Character Recognition,(IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (1) , 2016, 1-5
- Survey on Image Preprocessing Techniques to Improve OCR Accuracy <https://medium.com/technovators/survey-on-image-preprocessing-techniques-to-improve-ocr-accuracy-616ddb931b76>
- F. Simistira et al., "ICDAR2017 Competition on Layout Analysis for Challenging Medieval Manuscripts," 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, Japan, 2017, pp. 1361-1370, doi: 10.1109/ICDAR.2017.223.
- Clérice, Thibault. (2022). You Actually Look Twice At it (YALTAi): using an object detection approach instead of region segmentation within the Kraken engine. 10.48550/arXiv.2207.11230.
- Fizaine FC, Bard P, Paindavoine M, Robin C, Bouyé E, Lefèvre R, Vinter A. Historical Text Line Segmentation Using Deep Learning Algorithms: Mask-RCNN against U-Net Networks. Journal of Imaging. 2024; 10(3):65. <https://doi.org/10.3390/jimaging10030065>
- Leifert, Gundram, Christel Annemieke Romein, Achim Rabus, Phillip Benjamin Ströbel, Benjamin Kiessling, & Tobias Hödel. Evaluating State-of-the-art Handwritten Text Recognition (HTR) Engines; with Large Language Models (LLMs) for Historical Document Digitisation. Zenodo, 7 December 2023 r. <https://doi.org/10.5281/zenodo.8102666>
- Weidemann, M., Michael, J., Gruning, T., and Labahn, R. " (2018). HTR Engine Based on NNs P2 Building Deep Architectures with TensorFlow. Technical report.
- Peter Stokes, Benjamin Kiessling. Sharing Data for Handwritten Text Recognition (HTR). Digital Humanities in Practice, In press. fffhal-04444641f

Unread references (for the future)

- Petitpierre, R., Kramer, M. & Rappo, L. An end-to-end pipeline for historical censuses processing. *IJDAR* **26**, 419–432 (2023). <https://doi.org/10.1007/s10032-023-00428-9>
- Kass, Dmitrijs, and Ekta Vats. "AttentionHTR: Handwritten text recognition based on attention encoder-decoder networks." In *International Workshop on Document Analysis Systems*, pp. 507-522. Cham: Springer International Publishing, 2022.
- Ströbel, Phillip & Clematide, Simon & Volk, Martin & Schwitter, Raphael & Hodel, Tobias & Schoch, David. (2022). Evaluation of HTR models without Ground Truth Material.
- de Sousa Neto, A.F., Bezerra, B.L.D., de Moura, G.C.D. et al. Data Augmentation for Offline Handwritten Text Recognition: A Systematic Literature Review. *SN COMPUT. SCI.* **5**, 258 (2024). <https://doi.org/10.1007/s42979-023-02583-6>
- Ströbel, Phillip & Hodel, Tobias & Boente, Walter & Volk, Martin. (2023). The Adaptability of a Transformer-Based OCR Model for Historical Documents. [10.1007/978-3-031-41498-5_3](https://doi.org/10.1007/978-3-031-41498-5_3).
- Ströbel, Phillip & Clematide, Simon & Volk, Martin. (2020). How Much Data Do You Need? About the Creation of a Ground Truth for Black Letter and the Effectiveness of Neural OCR.
- AlKendi, Wissam, Franck Gechter, Laurent Heyberger, and Christophe Guyeux. 2024. "Advancements and Challenges in Handwritten Text Recognition: A Comprehensive Survey" *Journal of Imaging* 10, no. 1: 18. <https://doi.org/10.3390/jimaging10010018>

HTR: automatic transcription of handwritten texts in manuscripts (and early printed books)

- postal address interpretation,
- bank-cheque processing,
- signature verification,
- biometric writer identification,
- **manuscripts, documents, archives**



https://github.com/ai-forever/digital_peter_ajj2020

Still tough, but even for humans

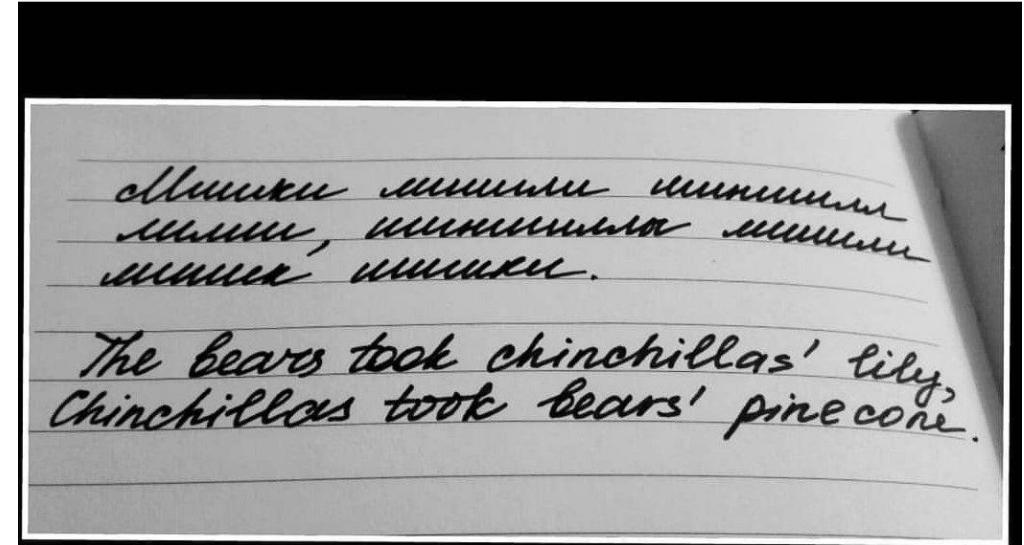




Unreadable by non-natives



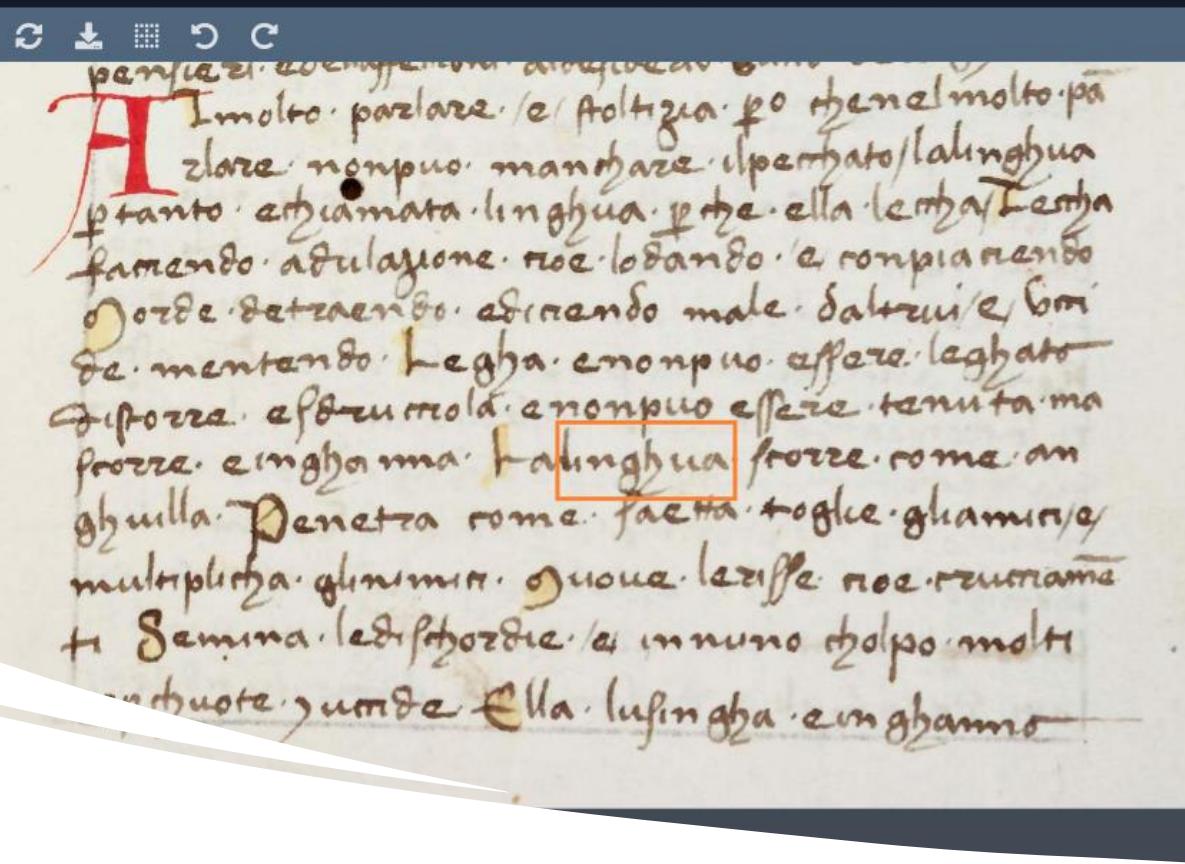
ailemira • [Follow](#)



Russian cursive

=

foreigners' nightmare :)



All < Previous Next >

Les dangers de la parole débridée

Tutorial Author: Clélia Nicolai Difficulty: easy Score: 27/99 (27%)

The screenshot shows a digital interface for transcribing medieval handwriting. At the top, there's a toolbar with icons for zoom, download, and other functions. Below it is a header with navigation links: 'All', '< Previous', 'Next >'. The main area is titled 'Les dangers de la parole débridée'. It displays the text from the manuscript in a grid format where each word is a separate box. Some words are colored green (e.g., 'Al', 'molto', 'parlare', 'famendo', 'adulazione', 'noe', 'lodando', 'ronpianendo', 'mentendo', 'Legha', 'affere', 'leghato', 'frorre', 'e in ghamma', 'falinghua', 'frore', 'come', 'ghilla', 'Peneta', 'facta', 'toglie', 'ghiamis', 'multiplica', 'glisimis', 'quoua', 'leriffe', 'noe', 'rurram', 'Semina', 'ledischorde', 'in nuno', 'cholpo', 'molti', 'achuote', 'jumde', 'Ella', 'lufingha', 'e in ghammo'). Other words are in red boxes (e.g., 'non', 'puo', 'mangiare', 'tanto', 'e', 'linghua', 'e', 'dinendo', 'male', 'e', 'Line 8, Part 5', 'puo', 'effere', 'tenuta', 'la', 'li'). There are also several empty white boxes. A progress bar at the bottom indicates 'Line 8, Part 5'.

Unreadable by non-experts

Test your paleographer skills: <https://www.multipal.fr>

HTR Pipeline

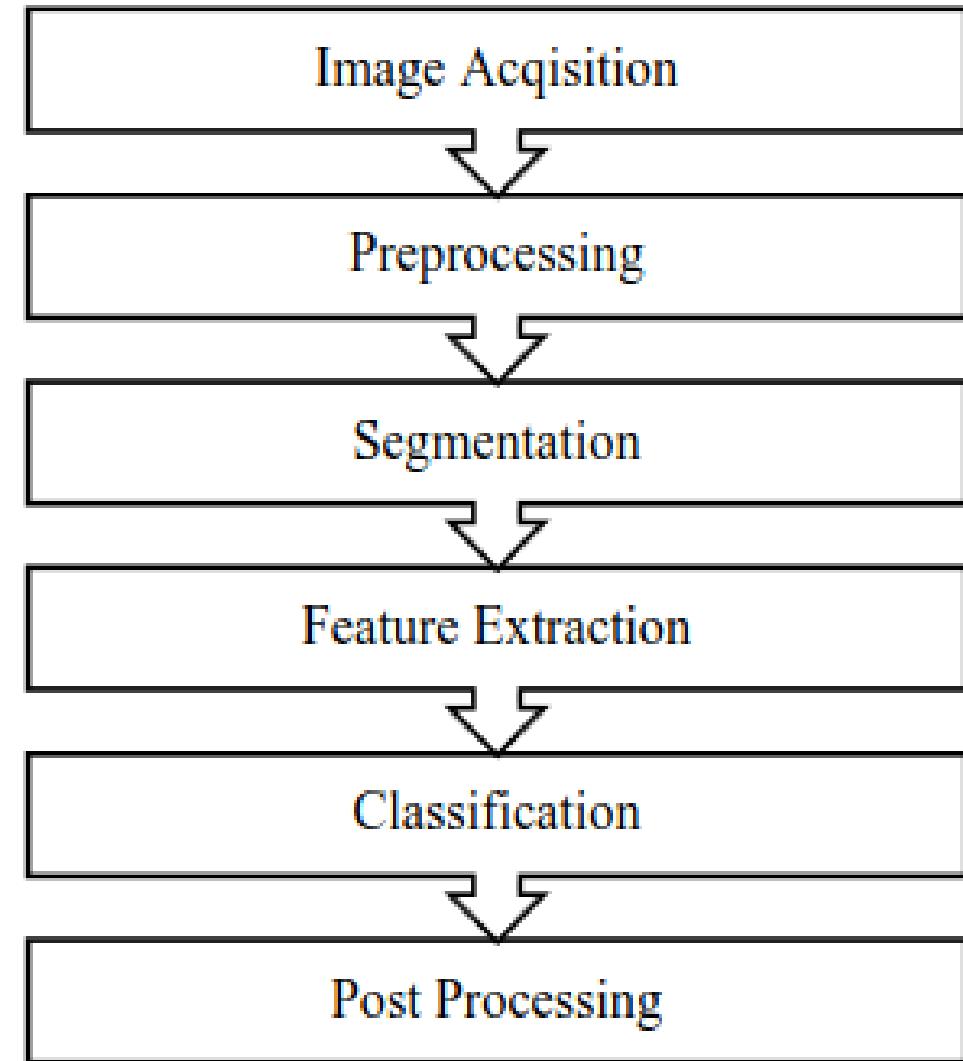
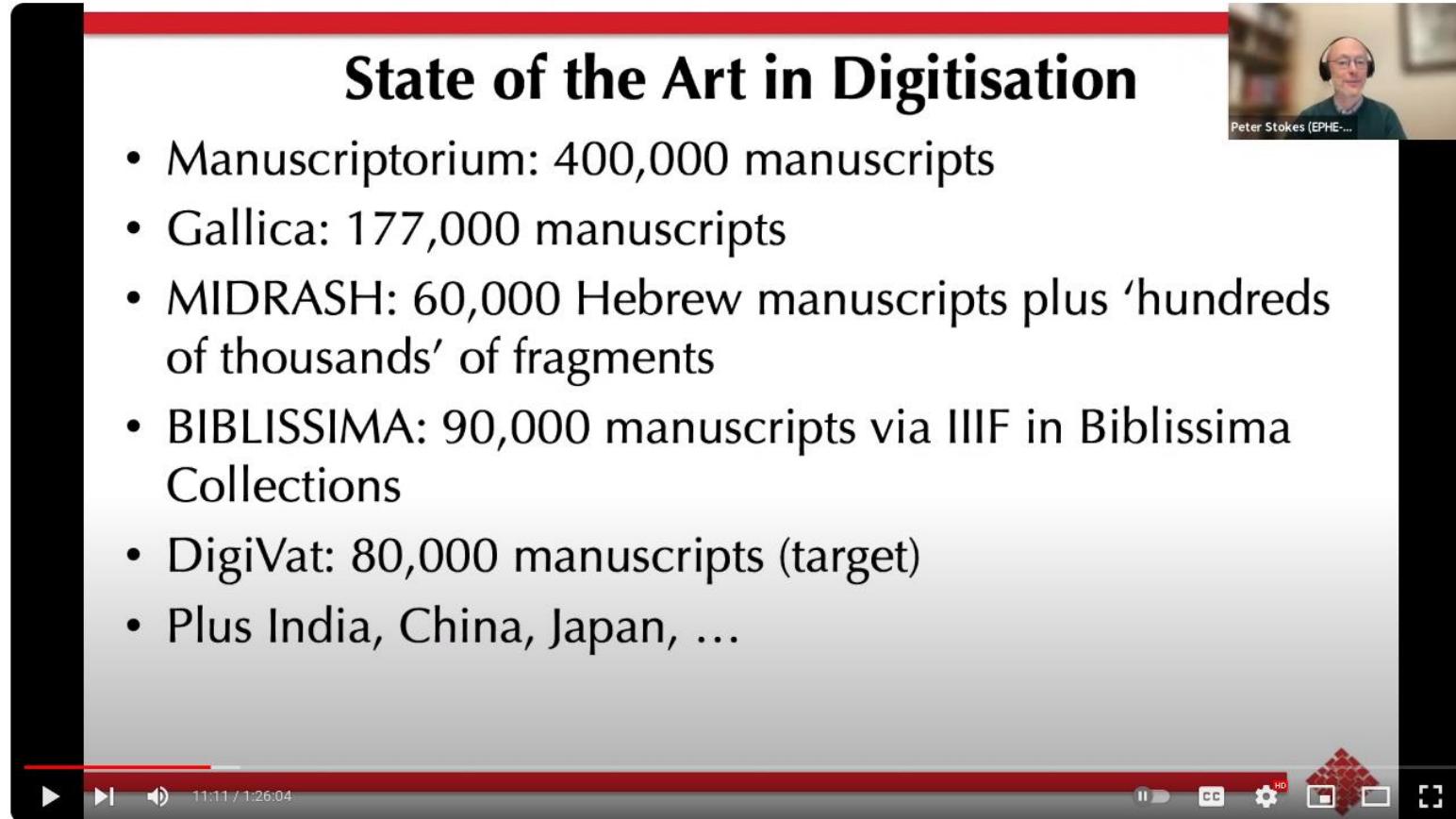


Image Acquisition

State of the Art in Digitisation

- Manuscriptorium: 400,000 manuscripts
- Gallica: 177,000 manuscripts
- MIDRASH: 60,000 Hebrew manuscripts plus 'hundreds of thousands' of fragments
- BIBLISSIMA: 90,000 manuscripts via IIIF in Biblissima Collections
- DigiVat: 80,000 manuscripts (target)
- Plus India, China, Japan, ...



Peter Stokes (EPHE-...)

How to Transcribe a Million Manuscripts with eScriptorium

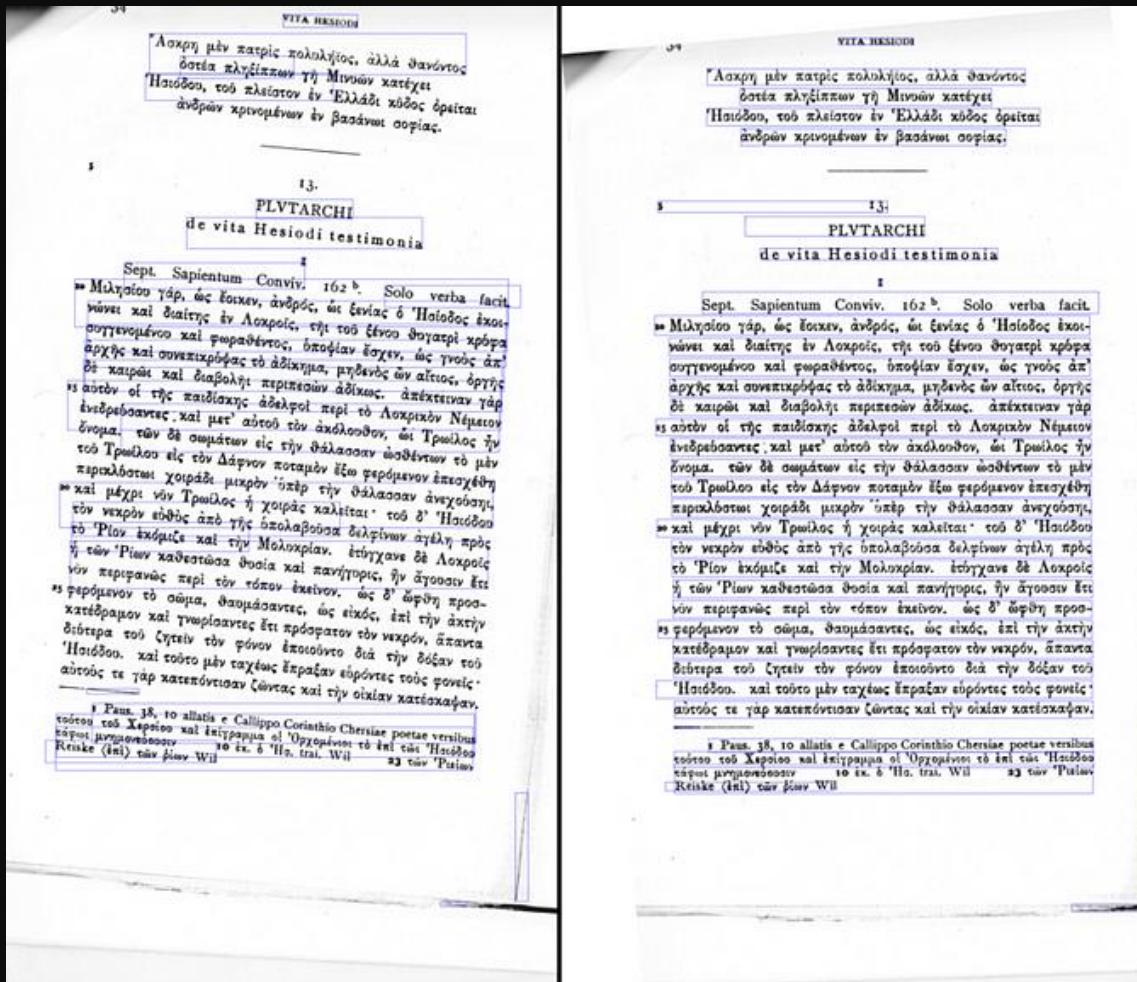
[How to Transcribe a Million Manuscripts with eScriptorium](#)

Preprocessing: Image Quality

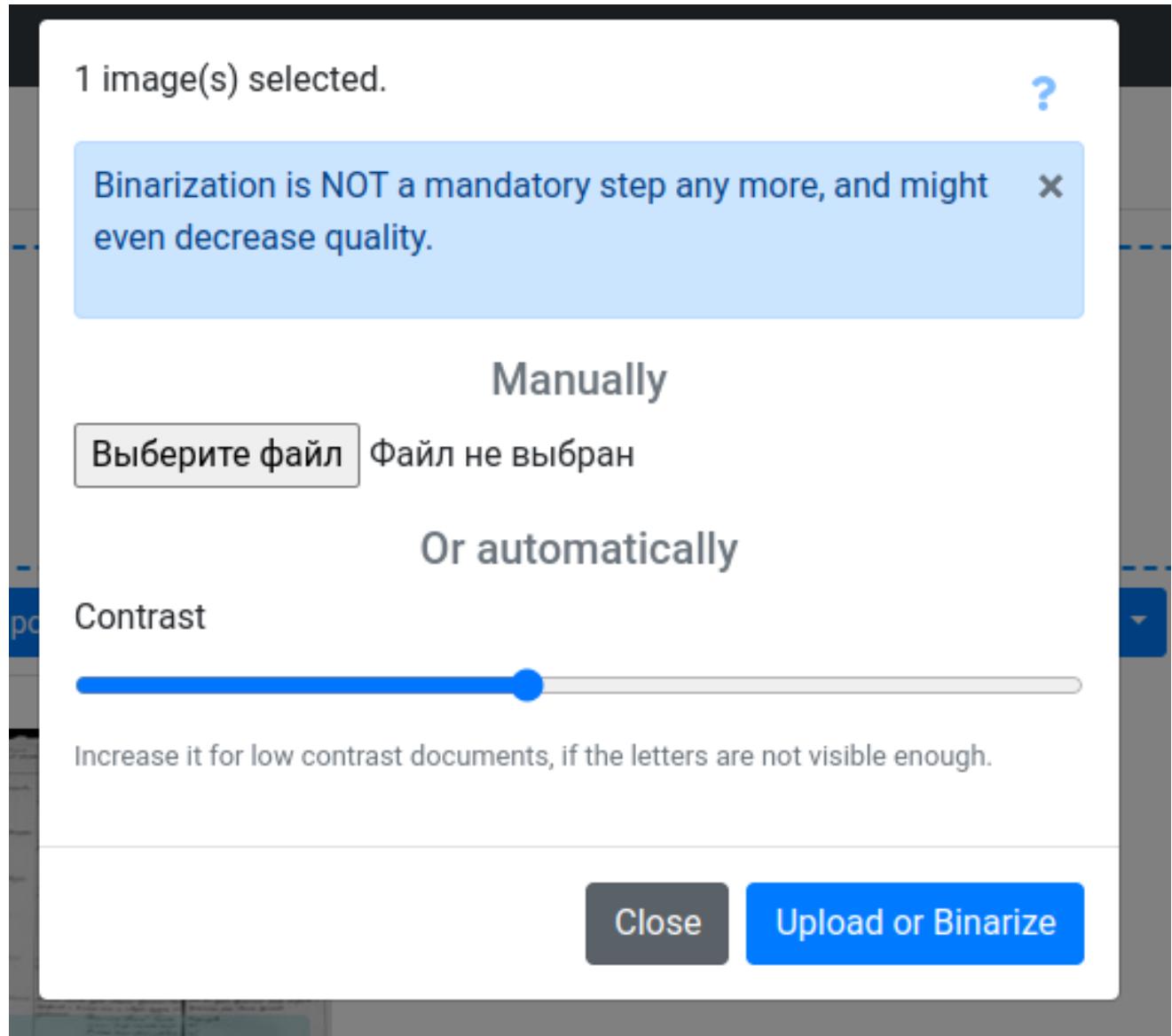
- Resolution
 - Contrast & Sharpness
 - Geometric transformations and so on

Survey on Image Preprocessing Techniques to Improve OCR Accuracy

<https://medium.com/techno-vators/survey-on-image-preprocessing-techniques-to-improve-ocr-accuracy-616ddb931b76>



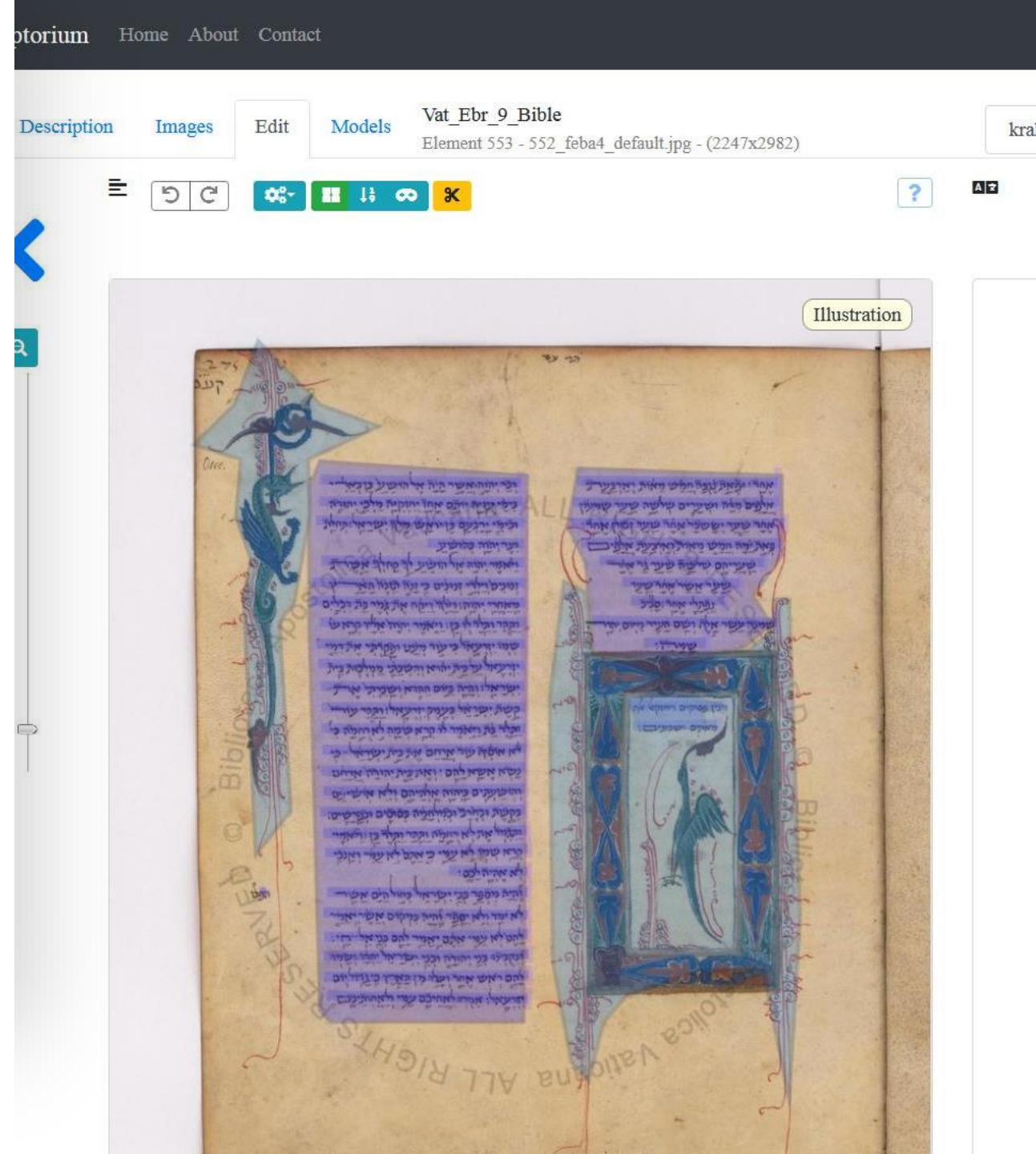
Preprocessing: Binarisation?



Segmentation

Why is it important?

- Reduces document complexity,
- Improves the accuracy of recognition algorithms,
- Is required for further analysis and automatic markup at the post-processing and publishing stages



Page number

62

Running header

ΣΟΦΟΚΛΕΟΥΣ

Primary text

XO. ἀνὴρ φρονεῖν ἔσικεν. ἀλλ' ἀνοίγετε·
τάχ' ἀν τιν' αἰδῶ καπτ' ἐμοὶ βλέψας λάβοι.

Line number

345

TE. ιδού, διοίγω· προσβλέπειν δὲ ἔξεστι σοι
τὰ τοῦδε πράγη, καῦτὸς ὡς ἔχων κυρεῖ.

AI. ίὼ

2 φίλοι ναυβάται, μόνοι ἐμῶν φίλων,
3 μόνοι ἐπ' ἐμμένοντες ὄρθῳ νόμῳ,
4 ἴδεσθέ μ' οἶον ἄρτι κῦμα φοινίας ὑπὸ ζάλης
5 ἀμφιδρομον κυκλεῖται.

Line number

350

XO. οἵμ' ὡς ἔσικας ὄρθᾳ μαρτυρεῖν ἄγαν.
δῆλοι δὲ τοῦργον ὡς ἀφροντίστας ἔχει.

Line number

355

AI. ίὼ

2 γένος ναῖτας ἀρωγὸν τέχνας,

App. Crit.

344 L has the *r* of *έσικεν* from a later hand.—*δρολγετε*] Wecklein writes *δρογε* δή.
345 *χδτ'*] made in L from *χδτ'*: this is explained by the false reading *χδτ'* ἐμοῖ
in Pal.—Blaydes conj. *καὶ* ἐμὲ βλέψας. **346** f. *ἴὼ..φίλων* is one

Commentary

Ajax, whose mother was Eriboea (569).—*τὸν εἰσαει.. χρόνον*: the phrase *τὸν δει χρόνον* is frequent. The separation of *τὸν εἰσαει* from *χρόνον* deserves notice, as suggesting the possibility that *τὸν δει* may sometimes have been used (without *χρόνον*) as = 'for ever': a usage which, however, lacks proof: see on *EI.* 1075.—*ληλατήσατο*: he had gone to the uplands of Mysia (730), 'in pursuit of foes' (564) to be despoiled. Cp. Thuc. 1. 11 § 1 (the Greeks at Troy) *φανορρα...πρὸς γεωργίας τῆς Χερσονήσου τραχύμενος καὶ ληστεῖας τῆς τροφῆς διώρει*. The *Iliad*

Commentary

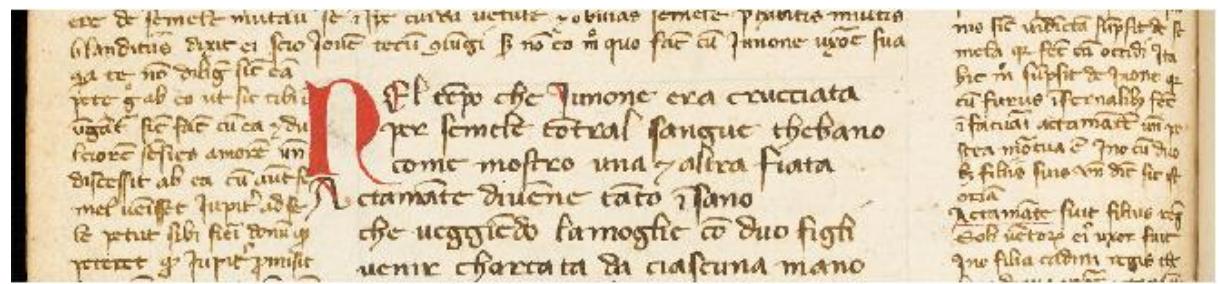
καὶτ' ἐμοῖ: for this modest *καὶ*, cp. *Pk.* 192 *εἴπερ καὶγά τι φρονῶ*, and n. on *Ant.* 719. *βλέψας* should naturally go with *καὶτ'* ἐμοῖ, though *βλέψει τοι τοι* seems to occur nowhere else, and *τριβλέψεις τοι* only in Lucian *Astr.* 20 *καὶ σφίσι γυανένεσι τῷ μὲν ἡ Ἀφροδίτη τῷ δὲ δὲ Ζεὺς...τριβλέψεις* ('looked with favour'). The alternative is to take *τῷ* ἐμοῖ as = 'in my case,' and *βλέψας* as epexegetic; but this is certainly harsh.

346 f. *διοίγω*: cp. *O. T.* 1287 *διοίγεις κλῆρος*, and *ib.* 1295.—*πράγμη*, deeds: cp. 21.

Najem-Meyer, Sven and Matteo Romanello. "Page Layout Analysis of Text-heavy Historical Documents: a Comparison of Textual and Visual Approaches." Workshop on Computational Humanities Research (2022).

Figure 2: The main layout elements of a scholarly commentary page.

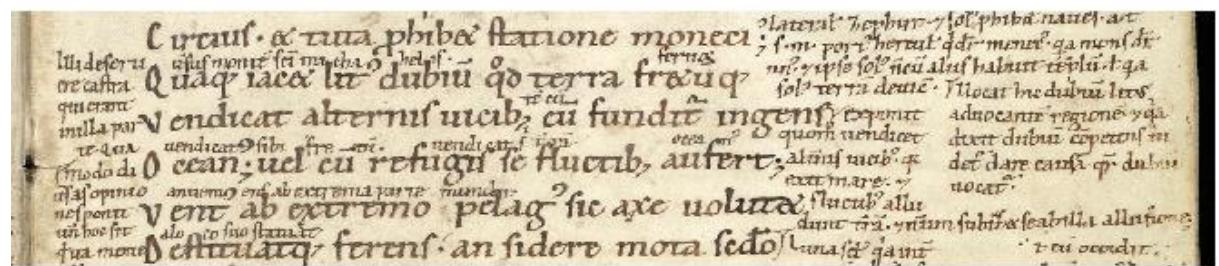
Most segmenters focus
on pixel classification



(a) CB55, p.67v



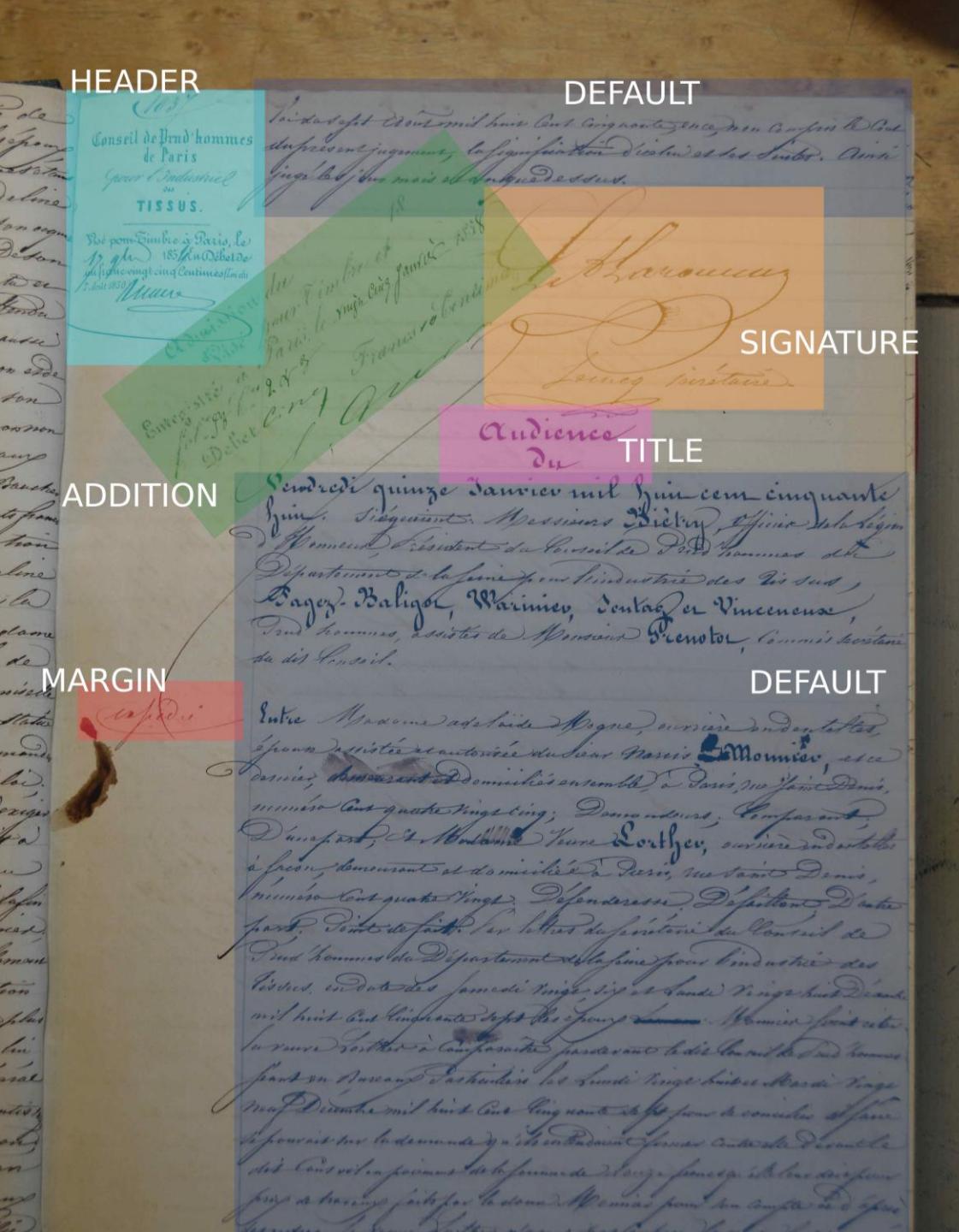
(b) CSG18, p.116



(c) CSG863, p.17

F. Simistira et al., "ICDAR2017 Competition on Layout Analysis for Challenging Medieval Manuscripts," 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, Japan, 2017, pp. 1361-1370, doi: 10.1109/ICDAR.2017.223.

Figure 1. Samples of pages of the three medieval manuscripts



HEADER

Conseil de Prud'hommes
de Paris
pour l'industriel
TISSUS.

Recd pour Timbie à Paris, le
11 juillet 1855. Au Débet de
plus ou moins cinq centimes francs
le dolt 1855. *Maurice*

ADDITION

MARGIN

expediti

DEFAULT

SIGNATURE

Audience
du **TITLE**

DEFAULT

Segmentation: SegmOntc

<https://segmonto.github.io/>

A Controlled Vocabulary to Describe the Layout of Pages

SegmOnto Region classes

... to make the published dataset as widely reusable as possible, we mapped our classes to the SegmOnto controlled vocabulary (Romanello, Najem-Meyer 2022)

Table 1

Complete list of fine- and coarse-grained page region classes used for layout annotation, with their corresponding mapping to SegmOnto's controlled vocabulary.

Fine	Coarse	SegmOnto Type:Subtype
commentary	commentary	MainZone:commentary
critical apparatus	critical apparatus	MarginTextZone:criticalApparatus
footnotes	footnotes	MarginTextZone:footnotes
page number	number	NumberingZone:pageNumber
text number	number	NumberingZone:textNumber
bibliography	others	MainZone:bibliography
handwritten marginalia	others	MarginTextZone:handwrittenNote
index	others	MainZone:index
others	others	CustomZone
printed marginalia	others	MarginTextZone:printedNote
table of contents	others	MainZone:ToC
title	others	TitlePageZone
translation	others	MainZone:translation
appendix	paratext	MainZone:appendix
introduction	paratext	MainZone:introduction
preface	paratext	MainZone:preface
primary text	primary text	MainZone:primaryText
running header	running header	RunningTitleZone

Zone & Lines types

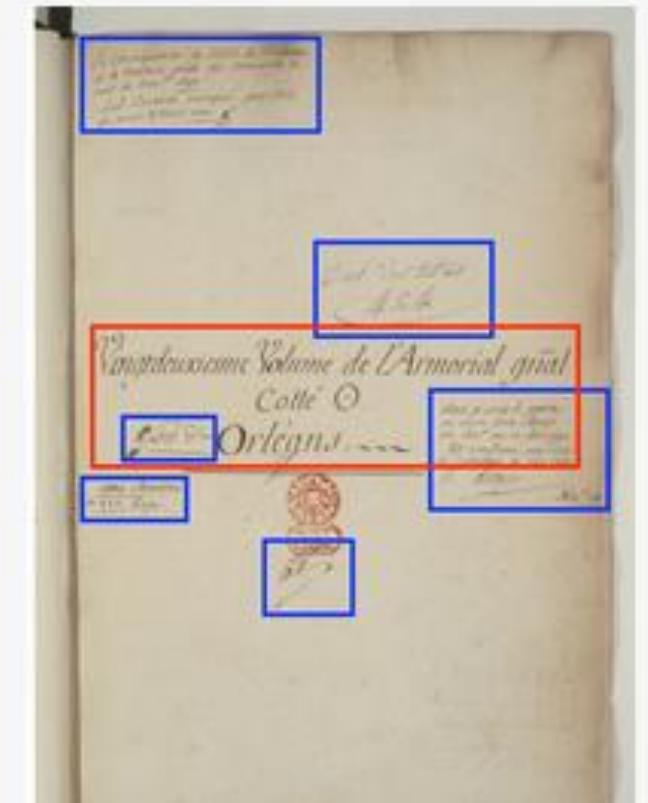
Why it is important?

MainZone:column#1

MainZone:column#2



TitlePageZone
and MarginTextZone





RunningTitleZone

Г о р о д а

15

На Исаакіевской площаdi.
По правой сторонѣ опъ Англійской набережной
до синяго мосша.

№ Правицельствующій Сенатъ.
203 Кусовникова, Николая, купца (бывшій Ус-
т'єва домъ).
Манежъ Лейбъ-гвардіи коннаго полка.
197 Бреммера, Христіана, бронзоваго масшера.
196 Нащокина, Александра Цепровича, Тайна-
го Совѣщника.
180 Нарышкина, Дмитрія Львовича, Оберъ-
Егермейшера.
179 Северина, Андрея, купца 1ї гильдіи.
160 Румплера золотыхъ дѣль масшера.

Большая морская.
Опъ Исаакіевской площаdi до Невскаго про-
спекта.

SegmOnto model

Layout analysis model trained
with [YALTAi](#), relying
on [YOLO](#) models, and [Kraken](#). Data
are annotated with
the [SegmOnto](#) controlled vocabulary.

<https://zenodo.org/records/10972956>

Published April 15, 2024

Image annotation tools

- **VGG Image Annotator (VIA)**
<https://www.robots.ox.ac.uk/~vgg/software/via/>
- **CVAT (Computer Vision Annotation Tool)**
<https://www.cvat.ai/>
- For manuscripts / books:
 - eScriotorium
 - Transkribus

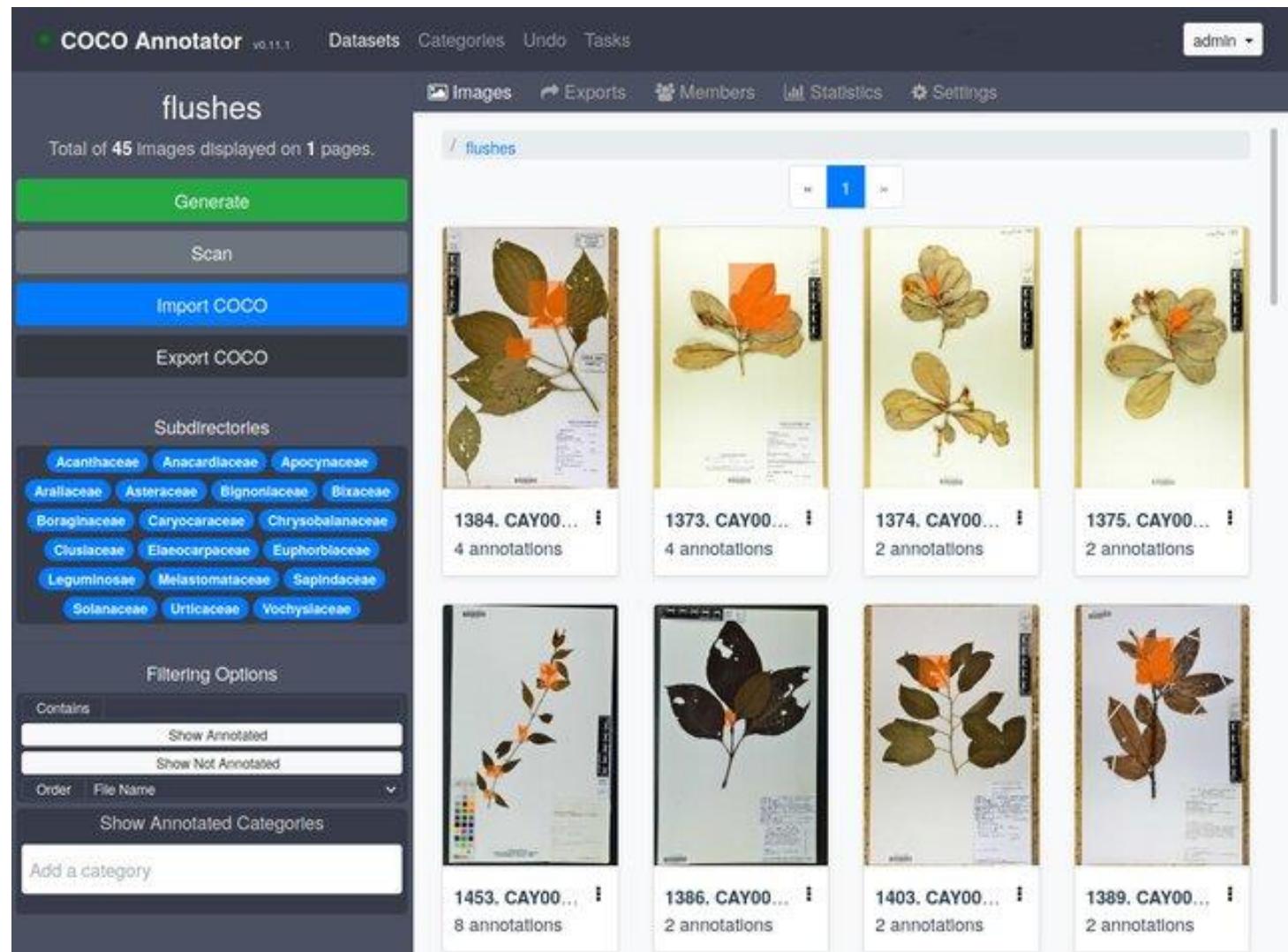


Image annotation

- Define an ontology of your segments and lines regarding your project goals
- Manually annotate (or use a model and then correct the result manually) a sample
- Train a model
- Check the output
- Repeat

Г О Р О Д А .		21
улицу; а что́лько чи́то на концѣ улицы прошивъ второго Адмиралтейства находишися домъ:		
No 213	На лѣвой сторонѣ.	No 231
Роговиковъхъ, Петра и Николая, купцовъ.	На правой сторонѣ.	Виаліе, Икона Васильевича, Айтиспинельного Спальского Собѣщника и Лейбъ-Медика.

214	Никифорова, Емельяна, купца.
215	Румянцова, Графа Николая Петровича, Министра иностранныхъ дѣлъ и Коммерціи.
216	Вара, Фомы, Англійскаго купца.
217	Фанзина, купца.
218	Криглера, Ивана, кузнеца.
219	Бергина, Хрисцифора, Коммерціи Совѣтника и Банкира.
220	Шульца, Ивана, пекаря.
221	Ейхера, Хрисциана, пекаря.
222	Пучковой, Анны Михайловны, жены Губернскаго Секретаря.
223	Моберлен, Ришарда Англійскаго купца.
224	Бардевика, Еремы Ивановича, купца 1-й гильдіи.
225	Бекели, Карла, купца 1-й гильдіи.

- 0 0.929952 0.064840 0.079710 0.025401
- 2 0.248792 0.233957 0.479469 0.144385
- 2 0.729469 0.235294 0.478261 0.145722
- 2 0.507246 0.643048 0.948068 0.617647
- 2 0.489130 0.117647 0.955314 0.056818
- 1 0.489130 0.062166 0.299517 0.036096

```

<Tags>
  <OtherTag ID="BT45" LABEL="NumberingZone:page" DESCRIPTION="block type NumberingZone:page"/>
  <OtherTag ID="BT46" LABEL="MainZone" DESCRIPTION="block type MainZone"/>
  <OtherTag ID="BT47" LABEL="RunningTitleZone" DESCRIPTION="block type RunningTitleZone"/>
  <OtherTag ID="BT48" LABEL="MainZone:column#1" DESCRIPTION="block type MainZone:column#1"/>
  <OtherTag ID="BT49" LABEL="MainZone:column#2" DESCRIPTION="block type MainZone:column#2"/>
</Tags>
<Layout>
  <Page WIDTH="621"
        HEIGHT="1122"
        PHYSICAL_IMG_NR="14"
        ID="eSc_dummypage_">
    <PrintSpace HPOS="0"
                VPOS="0"
                WIDTH="621"
                HEIGHT="1122">
        <TextBlock HPOS="304"
                  VPOS="187"
                  WIDTH="303"
                  HEIGHT="165"
                  ID="eSc_textblock_28141c0c"
                  TAGREFS="BT49">
            <Shape><Polygon POINTS="304 187 304 352 607 352 607 187"/></Shape>
        </TextBlock>
        <TextBlock HPOS="211"
                  VPOS="53"
                  WIDTH="188"
                  HEIGHT="34"
                  ID="eSc_textblock_33338d66"
                  TAGREFS="BT47">
            <Shape><Polygon POINTS="211 53 211 87 399 87 399 53"/></Shape>
        </TextBlock>
    </PrintSpace>
  </Page>
</Layout>

```

Training a model: YOLOv8

```
(venv) maria@maria-H410M-S2H:~/projects/spb$ yaltai alto-to-yolo export/*.xml my-dataset --shuffle .1 --segmonto region
Using list of inputs.
Found 20 to convert.
3/20 image for validation.
Shuffling data with a ratio of 0.1 for validation.
20it [00:00, 1246.01it/s]
20 ground truth XML files converted.
Configuration available at my-dataset/config.yml.
Label Map available at my-dataset/labelmap.txt.
Regions count:
- 00040 MainZone
- 00020 RunningTitleZone
- 00020 NumberingZone
(venv) maria@maria-H410M-S2H:~/projects/spb$ yolo task detect mode train model=yolov8n.pt data=my-dataset/config.yml epochs=100
```

You only look once (YOLO)

Train and Predict

```
Epoch    GPU_mem   box_loss   cls_loss   dfl_loss   Instances   Size
97/100    0G      1.159     1.81      1.233      6          960: 100% |██████████| 3/3 [00:06<00:00, 2.01s/it]
          Class Images Instances Box(P) R mAP50 mAP50-95: 100% |██████████| 1/1 [00:00<00:00, 2.18it/s]
          all      6       28      0.639     0.683      0.597     0.434

Epoch    GPU_mem   box_loss   cls_loss   dfl_loss   Instances   Size
98/100    0G      0.9454    1.63      1.281      7          960: 100% |██████████| 3/3 [00:05<00:00, 1.79s/it]
          Class Images Instances Box(P) R mAP50 mAP50-95: 100% |██████████| 1/1 [00:00<00:00, 1.98it/s]
          all      6       28      0.639     0.683      0.597     0.434

Epoch    GPU_mem   box_loss   cls_loss   dfl_loss   Instances   Size
99/100    0G      1.012     1.812     1.21       3          960: 100% |██████████| 3/3 [00:05<00:00, 1.99s/it]
          Class Images Instances Box(P) R mAP50 mAP50-95: 100% |██████████| 1/1 [00:00<00:00, 1.72it/s]
          all      6       28      0.639     0.683      0.597     0.434

Epoch    GPU_mem   box_loss   cls_loss   dfl_loss   Instances   Size
100/100   0G      0.9336    1.666     1.17       3          960: 100% |██████████| 3/3 [00:05<00:00, 1.97s/it]
          Class Images Instances Box(P) R mAP50 mAP50-95: 100% |██████████| 1/1 [00:00<00:00, 2.15it/s]
          all      6       28      0.632     0.683      0.596     0.43

100 epochs completed in 0.186 hours.
Optimizer stripped from runs/detect/train/weights/last.pt, 6.3MB
Optimizer stripped from runs/detect/train/weights/best.pt, 6.3MB

Validating runs/detect/train/weights/best.pt...
Ultraalytics YOLOv8.0.209 Python-3.10.12 torch-2.0.1+cu117 CPU (Intel Core(TM) i9-10900 2.80GHz)
Model summary (fused): 168 layers, 3006233 parameters, 0 gradients, 8.1 GFLOPs
          Class Images Instances Box(P) R mAP50 mAP50-95: 100% |██████████| 1/1 [00:00<00:00, 2.44it/s]
          all      6       28      0.642     0.683      0.599     0.458
          MainZone 6       12      0.502     0.75      0.533     0.469
          RunningTitleZone 6       6      0.943      1      0.995     0.78
          NumberingZone 6      10      0.48      0.3      0.268     0.124
Speed: 1.3ms preprocess, 49.6ms inference, 0.0ms loss, 7.1ms postprocess per image
Results saved to runs/detect/train
```

```
x1: tensor(12.9008, device='cuda:0') y1: tensor(105.7754, device='cuda:0') x2: tensor(609.1998, device='cuda:0') y2: tensor(1012.1533, device='cuda:0')
x1: tensor(194.3830, device='cuda:0') y1: tensor(50.6112, device='cuda:0') x2: tensor(425.4306, device='cuda:0') y2: tensor(88.9174, device='cuda:0')
x1: tensor(19.5257, device='cuda:0') y1: tensor(54.7710, device='cuda:0') x2: tensor(60.7622, device='cuda:0') y2: tensor(87.9373, device='cuda:0')
```

NumberingRunningTitleZone 0.77

Задание Описание MainZone 0.84

ЧЕТВЕРТАЯ АДМИРАЛТЕЙСКАЯ ЧАСТЬ.

Начинаясь отъ впечатенія мойки позади Галерного двора, идеть направо до крюкова и Никольского канала чрезъ фоншанку; содержаши въ себѣ обѣ Коломни и все проспранство, лежащее между лѣвымъ берегомъ фоншанки чрезъ городской ровъ и валъ, между заливомъ и Царскосельскимъ проспектомъ или Московскою дорогою до Лиговскаго канала и рѣчки Таракановки.

Въ сей часни находятся.

Три церкви:

- 1) Покровъ Божія матери близъ Аларчина моста, на Покровской улицѣ, 2) Св. Троицы близъ базаръ Измайловскаго полку и 3) Екатерины Мученицы при впечатеніи фоншанки въ Финской заливе, всѣ при деревянныи.

Площадь

Чашиной круглой рынокъ близъ Калинкина мосту.

Мосты:

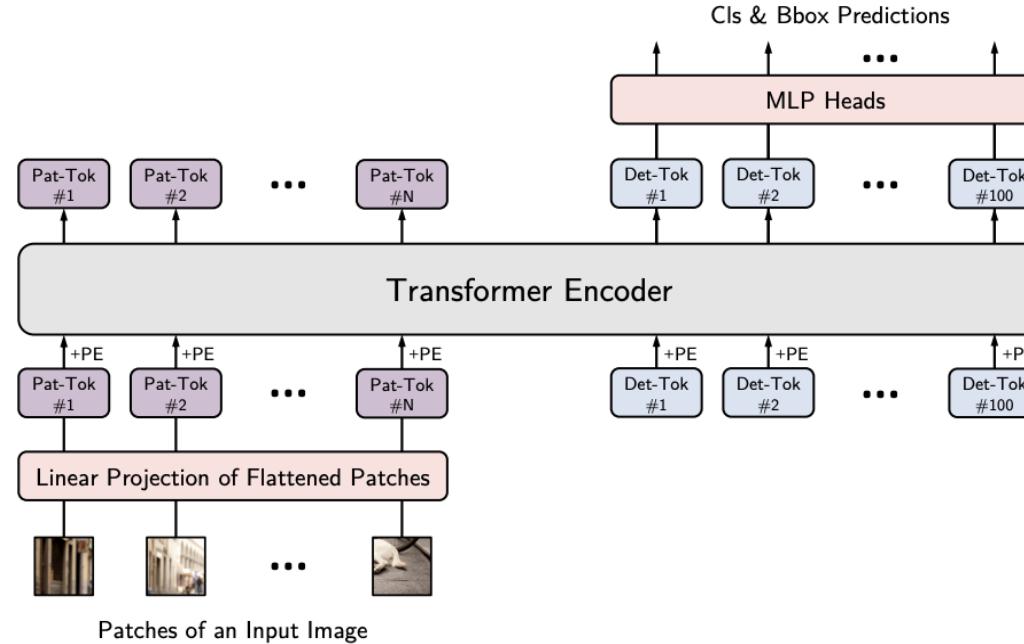
- a) Чрезъ мойку: Никольской, Крюковъ и Екатериненской каналы; и фоншанку

деревянные: 1) Канинь, 2) Торговой, 3) Офицерской, 4) Канальной, (границы со 2 Адм. частью) 5) Храповицкой, 6) Пикаловъ, 7) Сухарной, 8) Матисовской, 9) Банной, 10) Аларчинъ, (Границы съ Адм. частью), 11) Кузнецкой, 12) Николькой, 13) Малокалининской, 14) Крюковъ (Границы съ 3 Адм. частью).

Каменной чугунной мосинъ у Сальнаго буяна.

Каменные мосинъ чрезъ фоншанку: 1) Калинкинъ, 2) Памятниковъ, 3) Обуховъ, состоящи границы, съ 3 Адмир. и Москв. часью.

YOLO architecture for page layout

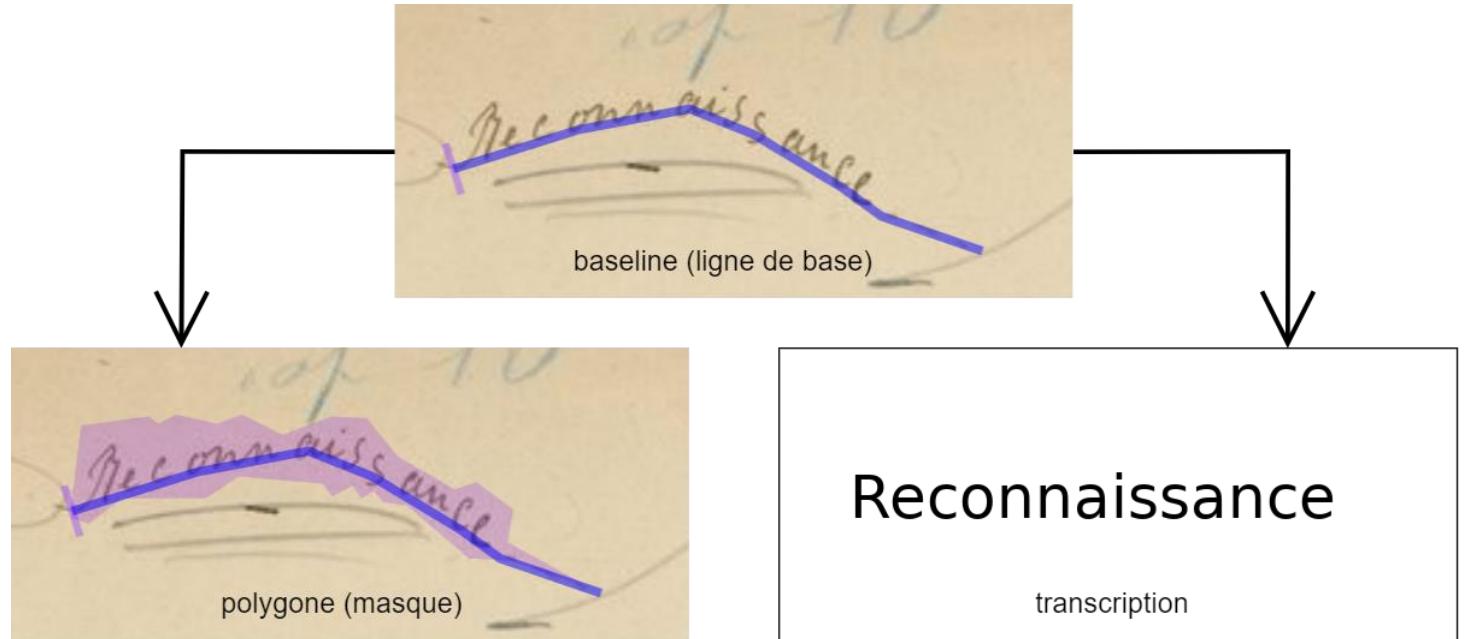


- From polygon (or bounding box) detection to pixel classification-based polygonization
- and now to object detection using isothetic rectangles

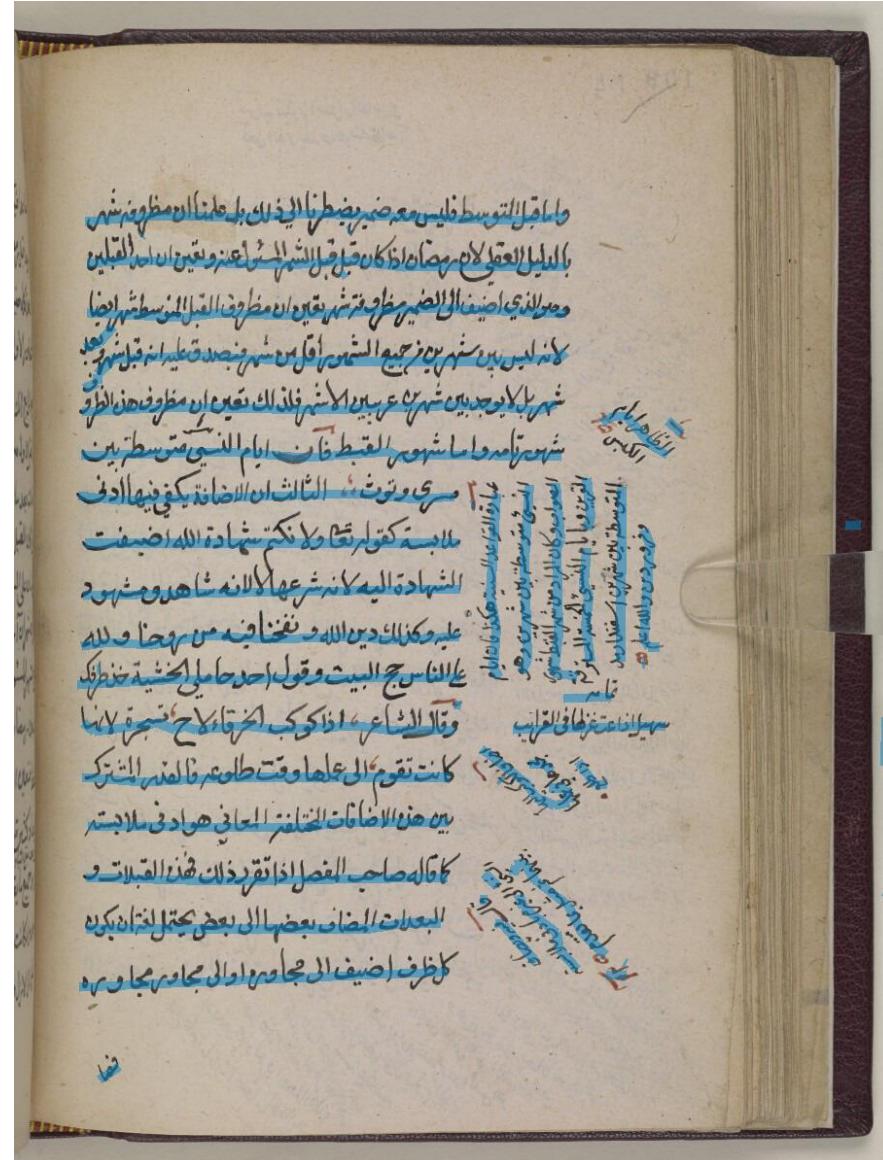
Clérice, Thibault. (2022). You Actually Look Twice At it (YALTAi): using an object detection approach instead of region segmentation within the Kraken engine. 10.48550/arXiv.2207.11230.

Baselines and Line Masks

- Use Geometric Analysis
- Apply Smoothing and Filtering
- Detect Local Minima/Maxima
- Fit Lines or Curves



Baselines

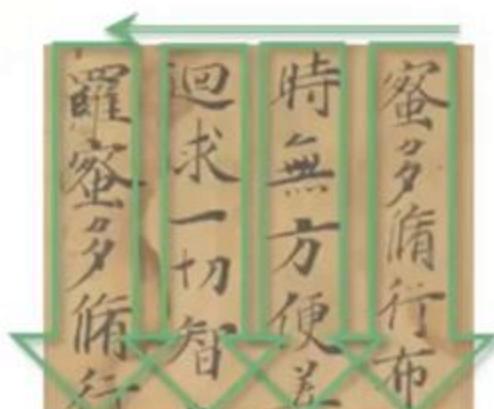




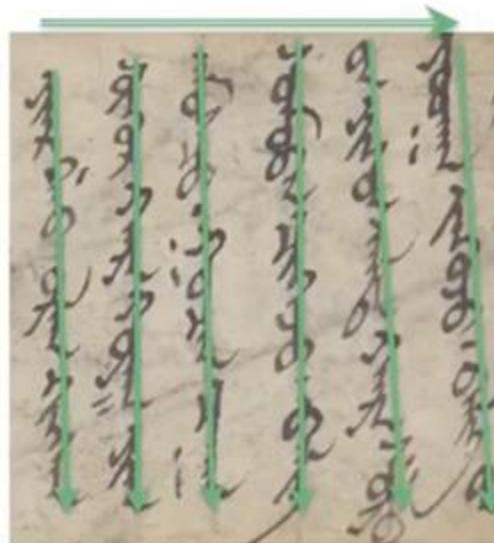
A 'simple' example: writing line & direction

topeum... dies caurus mutatq; numbro similis habbit. In genitivo singulari... tis. assumunt. Indati uo. r. amittunt et in e. conceptum

LtR then TtB, baseline



TtB then RtL,
L & R column lines



TtB then LtR,
vertical right 'base' line

אחד מהנערים אומר הפה
ראיתיכו ליש כי תלחמי
ידענו גבור חיל ואישם
מלחמות נכו זכר ואשת

RtL then TtB, topline



BtT, implicit
column lines



Clockwise,
internal baseline



- 01 OPEN HTR WITH ESCRIPTORIUM AND KRAKEN - Peter Stokes

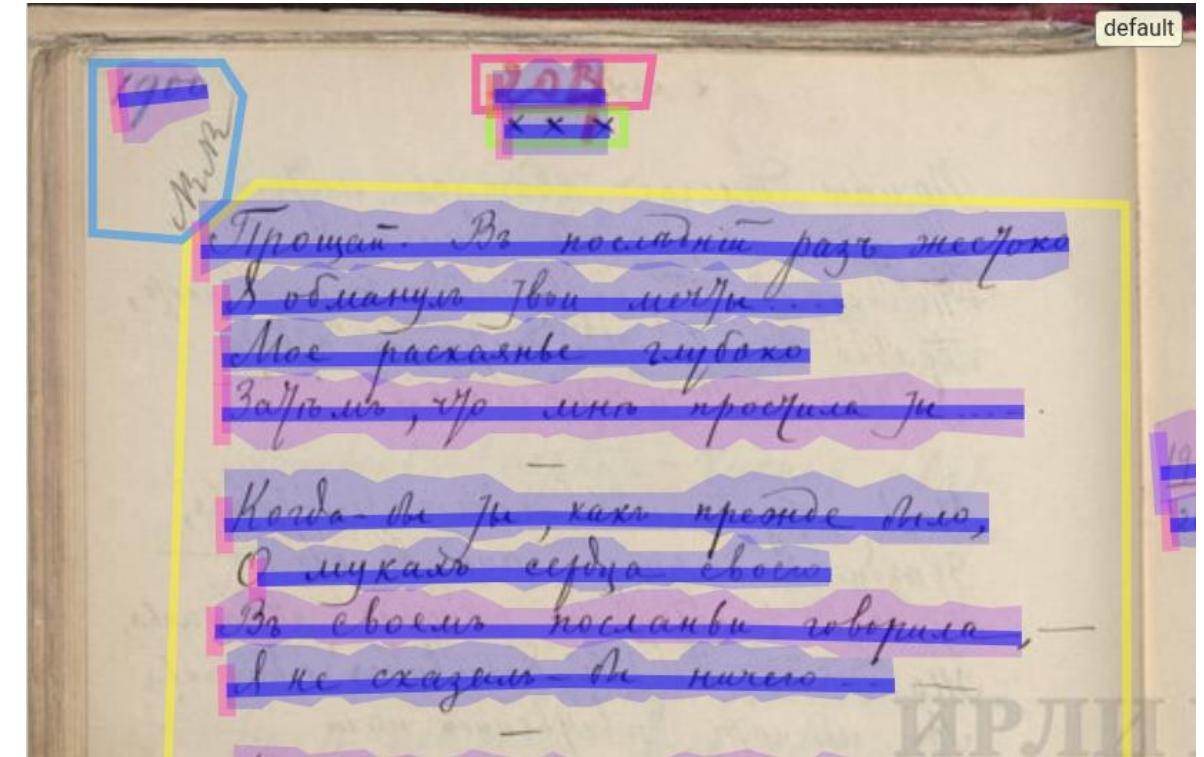


Line Masks

Fizaine FC, Bard P, Paindavoine M, Robin C, Bouyé E, Lefèvre R, Vinter A. Historical Text Line Segmentation Using Deep Learning Algorithms: Mask-RCNN against U-Net Networks. *Journal of Imaging*. 2024; 10(3):65. <https://doi.org/10.3390/jimaging10030065>

Look at the line masks

```
<TextBlock ID="eSc_dummyblock_>
    <String>
        WIDTH="251"
        HEIGHT="30"</String>
    </TextLine>
    <TextLine ID="eSc_line_ce789720"
        TAGREFS="LT6"
        BASELINE="113 186 454 185"
        HPOS="113"
        VPOS="168"
        WIDTH="341"
        HEIGHT="33">
        <Shape><Polygon POINTS="113 186 113 194 118 194 128 201 128 201 129 2
        <String CONTENT="Затѣм, что мнѣ простила ты...">
            HPOS="113"
            VPOS="168"
            WIDTH="341"
            HEIGHT="33"></String>
        </TextLine>
        <TextLine ID="eSc_line_b430f30e"
            TAGREFS="LT6"
            BASELINE="573 178 651 174 704 168 792 162 820 158 910 154"
            HPOS="572"
            VPOS="141"
            WIDTH="338"
            HEIGHT="46">
            <Shape><Polygon POINTS="730 156 730 156 729 156 727 157 710 158 708 1
            <String CONTENT="Мнѣ Твой, о милая, чертогъ...">
```



Cropped Line Masks

Затъм, что мнъ простила ты...

Рѣшить лазурныя загадки

Мнъ Твой, о милая, чертогъ...

Мое раскаянье глубоко

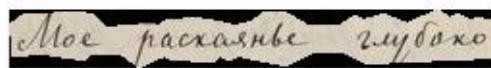
1900г.

HuggingFace models (TrOCR-base-ru)

⚡ Inference API ⓘ

Image-to-Text

Examples ▾



Computation time on cpu: 2.656 s

Мое раскаяны глубоко

«/» JSON Output

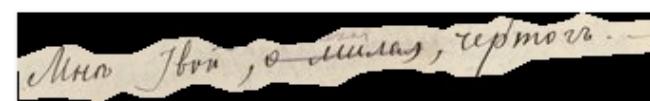
Maximize

Мое раскаянье глубоко

⚡ Inference API ⓘ

Image-to-Text

Examples ▾



Computation time on cpu: 3.751 s

Мнъ-Твой, милая, чертого...

«/» JSON Output

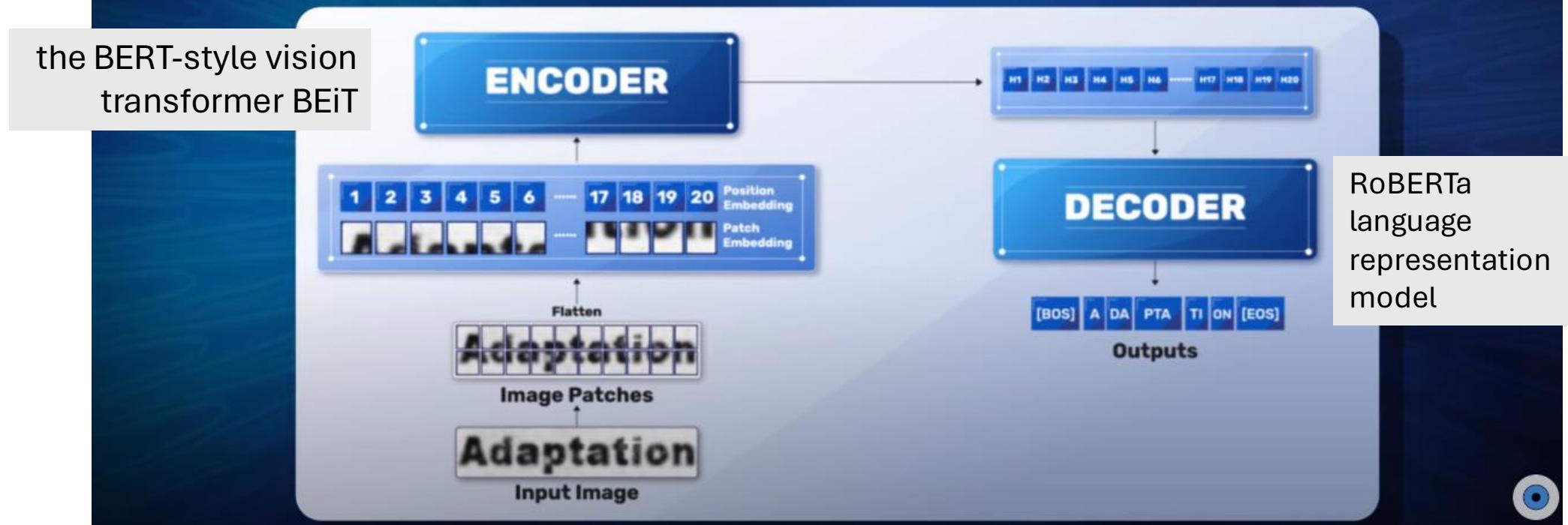
Maximize

Мнъ Твой, о милая, чертого...

TrOCR

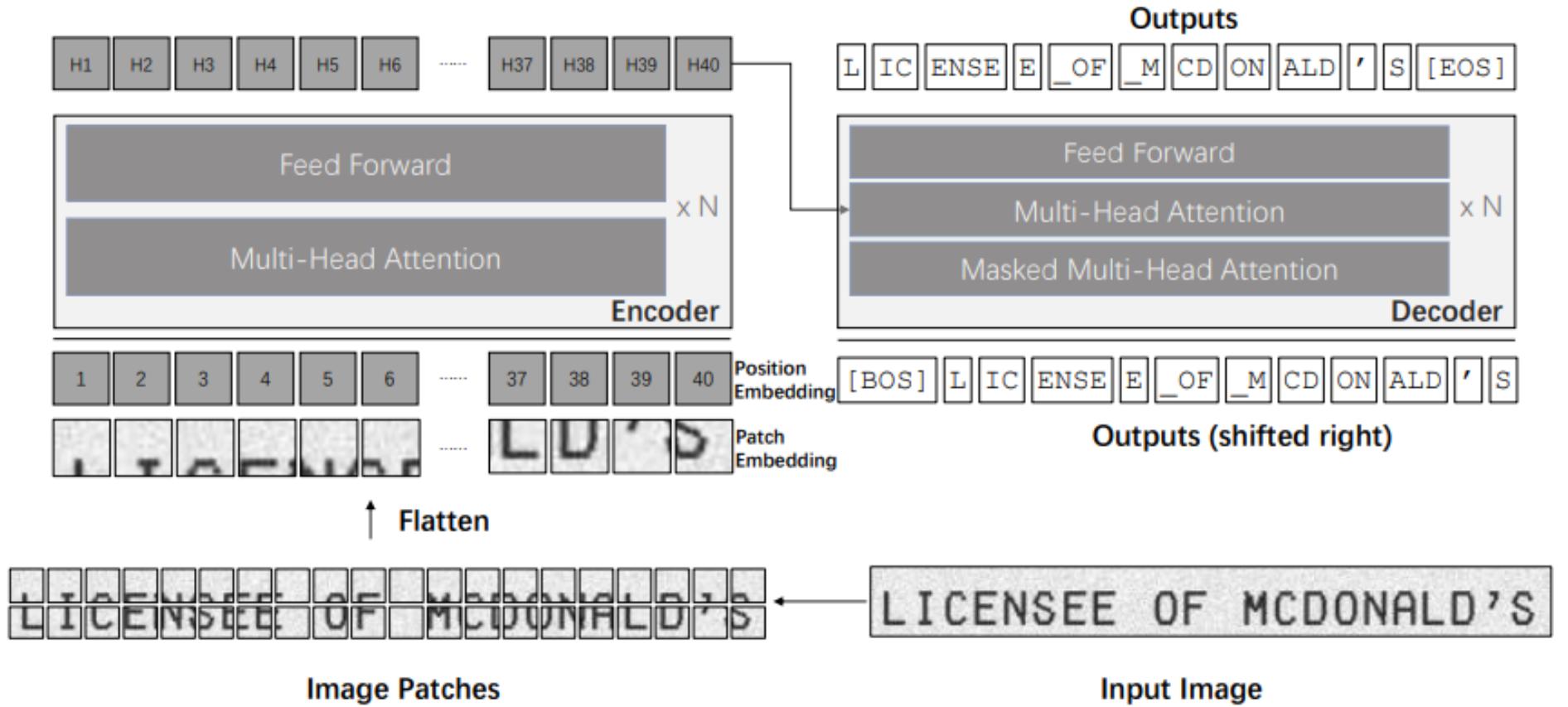
TrOCR Transformer Based OCR

the BERT-style vision
transformer BEiT



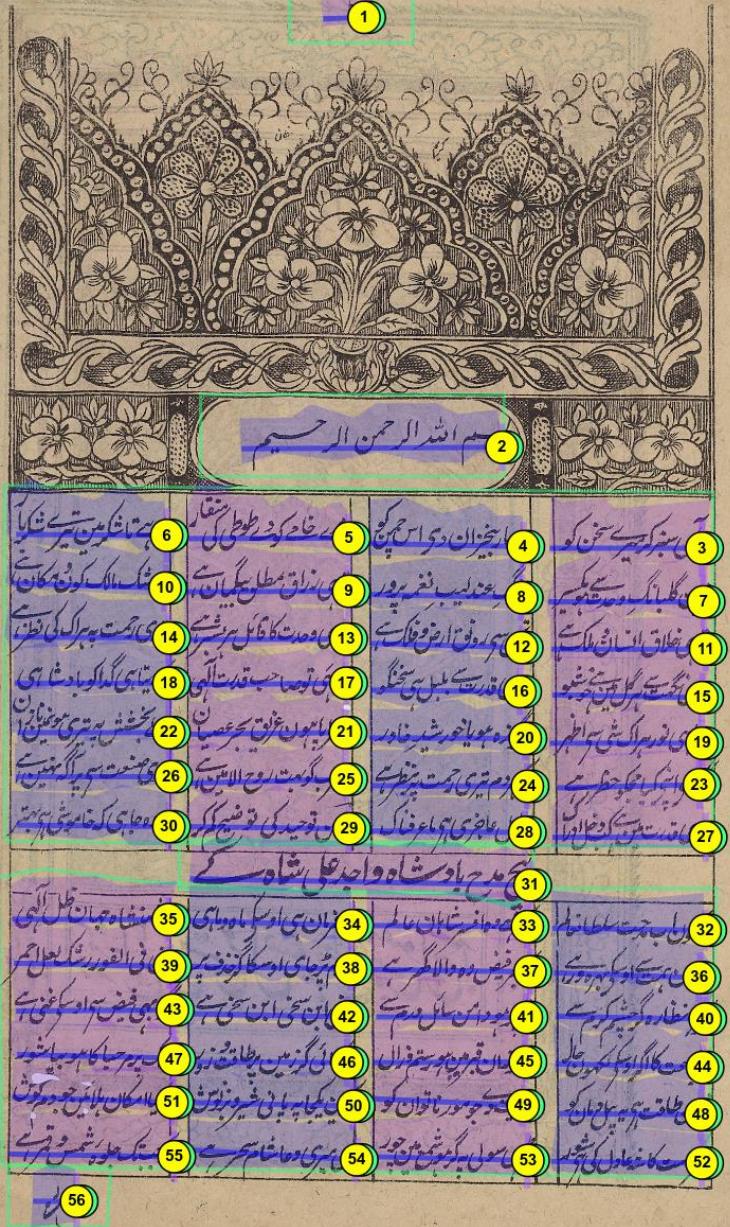
Li, Minghao, Tengchao Lv, Lei Cui, Yijuan Lu, Dinei A. F. Florêncio, Cha Zhang, Zhoujun Li and Furu Wei.
“TrOCR: Transformer-based Optical Character Recognition with Pre-trained Models.” *AAAI Conference on Artificial Intelligence* (2021).

TrOCR



combines the BERT-style vision transformer BEiT with a RoBERTa language representation model.

Ströbel, Phillip Benjamin, Simon Clematide, Martin Volk, and Tobias Hodel. "Transformer-based HTR for historical documents." *arXiv preprint arXiv:2203.11008* (2022).



نہ انتقالی حسن الرحمن الرحيم
بے شاکر میں نیرے شاکر
بلاشک مالک کوئی مکان کے
تو ابی زیاق مظلہ سیگیاں لے
تڑی رحمت پر اُنکی نظرم
کہ دینا بی گدا کو باشنا پے
سر ایاں عربی بھر عصیان
تڑی صنعت می پر اُنکی نظرم
چمن تو ہد کی تو ضیغ کم کر
دلیل عافری پے ماعرفات
بیچ می باشاد و اجد علی شاد کے
شکشاد جیان مل الی
تہ فرمان ابی اوسکی ماہ و مایس
پلی فیور رشک لعل حمر
قدم پنجاہی اوسکا گر خنف پر
کچی اب سخی این سخی لے
توبید بو دامن سائل ذرم می
تو لرزان قفر میں بو ستم درال
دکھانی گرزاں میں پ طلاقت زور
کریں پکھا پے پائی شیر و بنوس
چیزیں سول پے گر بوش من چور
کہ جب تک جلوہ نہیں و قمر لے

Reading order

Example of Baseline Segmentation
Singāsan Battīsī Nazm (1871)
Image: [Universitätsbibliothek Heidelberg](http://digi.ub.uni-heidelberg.de/diglit/caman1871/0002)

Metric: CER (character error rate)

Leifert, Gundram, Christel Annemieke Romein, Achim Rabus, Phillip Benjamin Ströbel, Benjamin Kiessling, и Tobias Hödel. Evaluating State-of-the-art Handwritten Text Recognition (HTR) Engines; with Large Language Models (llms) for Historical Document Digitisation. Zenodo, 7 December 2023 г. <https://doi.org/10.5281/zenodo.8102666>

$$\text{CER} = \frac{\sum_{i=1}^n \text{dist}_c(\text{pred}_i, \text{true}_i)}{\sum_{i=1}^n \text{len}_c(\text{true}_i)}$$

CER - base

Dataset	Roman Type Print		Republic (7)		Glagolitic		Shorthand	
Engine	w/o LM	w/ LM	w/o LM	w/ LM	w/o LM	w/ LM	w/o LM	w/ LM
HTR+	0.52%		0.62%			5.65%		13.63%
Pylaia	2.02%		2.98%		6.32%	6.05%	11.91%	11.18%
IDA	2.74%	2.65%	2.69%	2.16%	4.83%	4.54%	9.53%	9.29%
TrOCR	4.1%		3.24%		4.67%		9.70%	

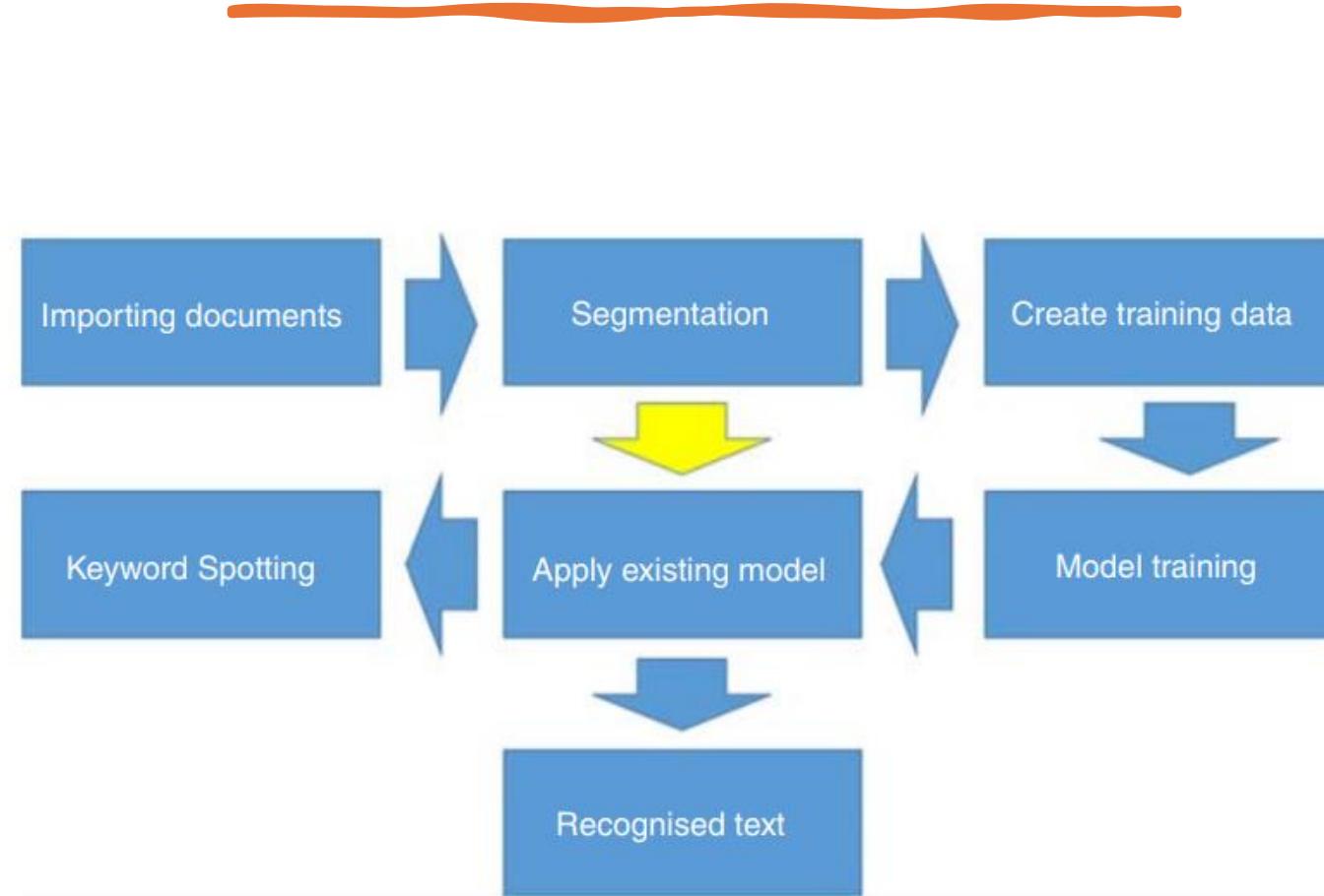
Transkribus

- **Archivists,**
- **humanities scholars,**
- **members of the public,**
- all of whom are interested in the study and exploitation of historical documents**

The screenshot shows the Transkribus homepage. At the top, there is a navigation bar with the Transkribus logo, links for "Platform & Features", "Solutions", "Resources", "The co-op", "Plans & pricing" (with a "New" badge), and buttons for "Open app", "Try free", and "EN". The main content area features a large image of a historical document with handwritten text. Overlaid on this image is a white rectangular box containing the text: "Publish your documents online with Transkribus Sites". Below this, a subtext reads: "Create a searchable, online database of your documents which can be accessed by anyone, from anywhere." Two buttons are present: "See features" (white background) and "Try for free" (dark blue background). To the right of the main image, a vertical sidebar displays a snippet of the historical text with some words highlighted in red, followed by a block of Dutch text explaining the context of the document.

van der groe nots, etc in presente van de nabesz. get.
my ten versoeke van dhr. antonio Bierens Coopman
binnen deser steede gevonden ende getransportert by ende
nevens den persoon van dhr. Adriaen Sewer mede
Coopman alhier aan ende aen den selven verschot prompte
ende datel betal van sekere orijuele wicels, een
d'hr. paul de fucks gehenne naer voor syn Cheur
vorstel doorluchticheit van brandenburgh in
Berlin gedirigiert ende hier naer van Woort tot
woort gecopieerd, inde met addres ten huyse van
de voors. hr. verwer, maer want deseive hr.
verweter antwoerde gaende seyde dat zy de voors.
Wissels niet soude bet. pr route van orde
soo hebbink nots. Voort nominique supra
wel expressel. geprotoest. van nondetal. mits=+
gaders van wissel ende herwijsel ende voorts van
alle Costen schaden ende Interessen alredie gehad.
gedien ende geleiden ende noch te hebben doen ende lyden
omme alle deselve te verslaen daer ende soon den
reqt. te radie werden sal alles opecht gedaen
t amst. ter presente van Jeremias vliet ende
ab. Vroewen als get.

Transkribus workflow



HTR+

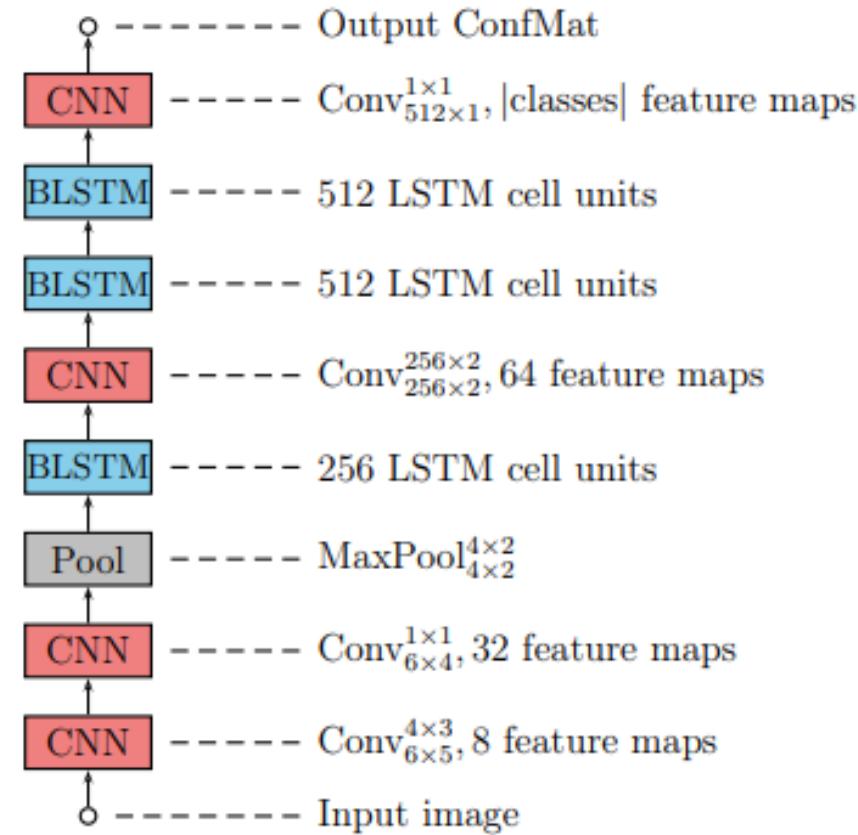


Figure 4: The proposed HTR model: A combination of various CNN and BLSTM blocks with one additional pooling layer. The subscript in $Conv/MaxPool_{m \times n}^{m \times n}$ describes the size of the kernel and the superscript defines the stride along both dimensions. Basically, $|classes|$ is the number of characters present in the ConfMat.

Transkribus "Smart" Models

- Resolve abbreviations
- Bring superscripts down to the line
- Add missing punctuation or graphemes
- Modernize/Unify the orthography
- Transcribe into a different alphabet

Smart models are for non-philological use!

- Quantitative studies
- Wider audience

Correct resolution of abbreviations

2-18 iže taku krêpostb podaeš
2-19 i s(ve)timъ m(u)č(e)n(i)komъ da da si n(a)mъ
2-20 onu tvoû m(i)l(o)stъ . i tvoe bl(agoslovle)n
2-21 ie . siê vsa b(og)u sl(a)vna . plna
2-22 g(lago)lb i pêniê . i ta sl(a)vitb t

APPLICATIONS FOR EDUCATIONAL INSTITUTIONS

NAME OF INSTITUTION	TOWN	AMOUNT	OBJECT	DATE	
Acadia University, (1)	P Wolfville, N.S.		General	1/14/11	Has done a lot
Americus Institute	Americus, Ga.	10000	Building	1/17/11	D + Low
Albemarle Normal & Ind. Inst.	Albemarle, N.C.		General	2/4/11	Low
Asbury College	Wilmore, Ky.		Buildings & Industrial Plant	2/13/11	D
Adrian College	Adrian, Mich.		Endowment	3/23/11	Denominational
Alabama State Normal School	Hopkinsville, Ala.	25000	Building	3/10/11	State interest
Antioch College	Yellow Springs, O.	100000	Endowment	3/22/11	Not sufficient
American Church Inst. for Negroes	New York, N.Y.		General	4/10/11	D
Amherst College	Amherst, Mass.	50000	Increase Salaries	4/10/11	Has done a lot
Alma College	Alma, Mich.		Library Building	5/18/11	Denominational
American International College	Springfield, Mass.		General	6/6/11	Not sufficient
Alberta Ladies' College (1)	Red Deer, Alta.		General	2/15/11	Not sufficient

Table and Field models (beta)

Allen University	Columbia, S.C.		Library Building	4/26/12	Low
Acadia University, (2)	P Wolfville, N.S.	25000	Library Building	5/2/12	Has done a lot
Abingdon Presbytery	Drapers, Va.	500	Building	5/14/12	D + Low
Amity College	College Springs, Ia.		Endowment	5/31/12	Not sufficient
Albert Lea College	Albert Lea, Minn.		Endowment	6/18/12	Denominational
Anderson College (1)	Anderson, S.C.		General	8/20/12	D + Low
Alabama University, Jr.	University, Ala.		Memorial Building	9/24/12	State interest

Data is all you need

- For the text written in one hand:
 - 15,000 transcribed words (\approx 75 pages).
- For the printed text:
 - 5,000 transcribed words (\approx 25 pages).

	Character Error Rate	Training Data	Approximate time to create the training data
Printed text	0,5-2%	~ 5.000 words / 25 pages	15 hours
Single hand - simple writing	2-4%	+10.000 words / 50 pages	25 hours
Several hands – all seen during training	4-6%	several 10.000 words training data / several 50 pages	25 hours x each hand
Many hands from same period and region – not all seen during training	6-8%	+100.000 words / more than 500 pages	~ 250 hours

~~HANZ. B. IT. I, c. I~~

1



Li sposi promessi

3

2

Capitolo I.

4

Quel ramo del lago di Como, che volge a
mezzogiorno, chiuso e come guidato da due
catene now interrotte di monti, stenden-
dosi in seni e golfi d'ineguale grandezza,
a seconda dello sporgere e del rientrare
di quelli, viene quasi tutto ad un tratto a
ristringeri ~~verso questo convegno~~
di fronte tra una montagna, ed un' an-
gusta missione formata soltamente dal

1 Nurte

2 14

3 N Spose promesse

4 Capitolo

5 Quel ramo del lago di Como, che volge a

6 mezzo giorno, chiuso, e come guidato da due

7 catene non interrotte di monti, stenden-

8 dosi in seni e golfi d'incuale grandezza

9 a seconda dello sporgere e del rientrare

10 di quelli, viene quasi tutto ad un tratto a

11 ristrengeri

12 a una montagna ed un am-

13 sia riviera formata sentamente dal

14 Deposito di tre grossi, e vicini torrenti,

15 Il lungo ponte, che in quel luogo con-

16 Cepende qui corpo et aspetto di fine

17 giunge le due rive, vende ancor più senti-

18 bile all'occhio questa trasformazione e par-

19 che divida il lago dall'Alda. A diritta la

scription

Ontology

Images

Edit

Models

Reports

I promessi sposi

 Drop images here or click to upload.

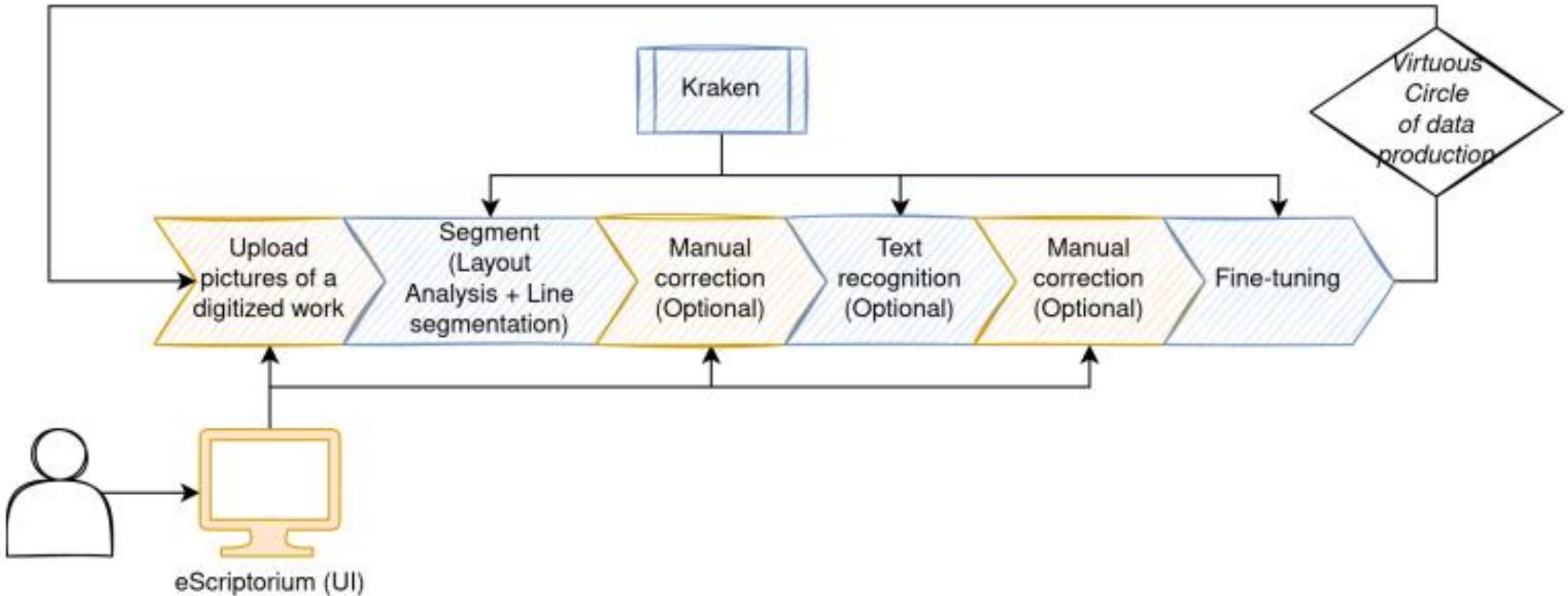
Unselect all Selected 0/48

Import Export Train Binarize Segm

1 2 3 4 5 6 7 8

eScriptorium
Open Source Software





Clérice, Thibault. (2022). You Actually Look Twice At it (YALTAi): using an object detection approach instead of region segmentation within the Kraken engine. 10.48550/arXiv.2207.11230.

Transkribus vs. eScriptorium



• Service for a broad audience, including researchers, historians, and archival enthusiasts.	Software built and supported by the community specifically towards digital humanities specialists, scholars, and researchers
• Online usage	Requires local installation, which may be challenging
• Commercial product with a free credits	Open Source
• Unable to export models	Able to import and export models
• Exports to Txt, Page xml, ALTO, PDF, TEI, docx	Exports to Txt, Page xml, and ALTO formats
• HTR+, PyLaia, TrOCR models	Kraken models

Sharing the data

State of the Art in Digitisation



- Manuscriptorium: 400,000 manuscripts
- Gallica: 177,000 manuscripts
- MIDRASH: 60,000 Hebrew manuscripts plus 'hundreds of thousands' of fragments
- BIBLISSIMA: 90,000 manuscripts via IIIF in Biblissima Collections
- DigiVat: 80,000 manuscripts (target)
- Plus India, China, Japan, ...

How to Transcribe a Million Manuscripts with eScriptorium

Kraken models for eScriptorium

- https://zenodo.org/communities/ocr_models/
- <https://zenodo.org/communities/scriptnet>

- Peter Stokes, Benjamin Kiessling. Sharing Data for Handwritten Text Recognition (HTR). Digital Humanities in Practice, In press. fffhal-04444641f
- [How to Transcribe a Million Manuscripts with eScriptorium](#)

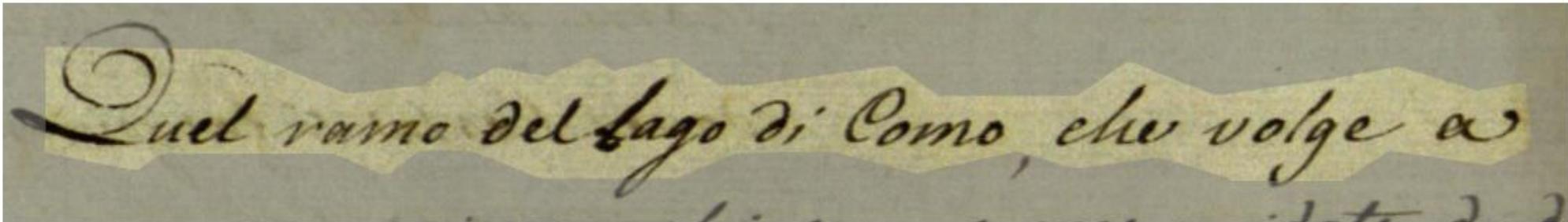
Home Contact My Projects My M

Line #4 X

Ontolo

WANT B

ssi



Duel ramo del Lago di Como ele volge a

by admin (kraken:Tridis_Medieval_EarlyModern) on Mon Apr 22 2024 11:26:11 GMT+0100

-Toggle transcription comparison ?

Duel ramo del lago di Como chele volge a
DQuel ramoe deul Lagoávo dio CoGmo eoleu volgse a&
Duel ramo del Lago di Como ele volge a

kraken:ManuMcFondue
kraken:Gallicorporat_best
kraken:Tridis_Medieval_EarlyModern (current)

*a seconda dello sporgere e dei rientrare
di quelli viene quasi tutto ad un tratto a
rstringersi. Dopo questo modo cede a grotta
di fiume tra una montagna, ed un' al-
tra riviera formata lentamente dal*

*a seconda dello sporgere e dei
ricontrarsi di quelli viene quasi tutto ad un
trattto a rstringersi. Dopo questo modo cede a grotta
di fiume tra una montagna, ed un'altra
riviera formata lentamente dal*

Google Document AI

Build your own Document OCR processor [Go to Document AI](#)

FIELDS	KEY VALUE PAIRS	TABLES	OCR
MANZ. B.it.I, C.I			
BIB			
Sti Sposi promessi			
Capitolo I			
Quel rame del Lago di Como, che volge			
Quel rame del Lago di Como, che volge			
Quel rame del Lago di Como, che volge			
mezzogiorno, chiuso, e come guidato da due			
e			
catene now interrotte di monti, stendesi			
dosi in seni e golfi d'ineguale grandezza,			
a seconda dello sporgere e del rientrare.			
di quelli, viene quasi tutto ad un tratto a			
vistringersi ad epito			
tra una montagna, ed un' am-			
pie riviera formata lentamente dal			
deposito di tre grossi, e vicini torrenti;			
le prende quici corso ed alpetto di fiume. I lungo sonte, che in quel luogo con-			
giunge le due rive, vende ancor più sensi-			
:ibile all'occhio questa trasformazione e par-			
che dividia il lago dall'Adda. A diritta, la			
testa del ponte passa su le radici del Monte			
Sauvachile; l'altra è piantata nel lembo			
della riva, che rende con lento pendio,			
appoggiata alle falde della montagna,			
assunse il Rojonone dai molti suoi			
congiunti acuti, e separati, a gura de-			
l'acqua tagli. Il fondo estremo, interrotto			
dallo scorrere dei torrenti, e di nulla e grilla			
ff			
monte			



You



Can you read the text from this image?

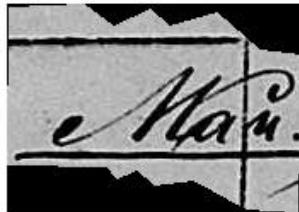


ChatGPT

The text in the image reads "Рождения." and "Крещения.", which in English translates to "Births." and "Baptisms." respectively.



You



And this?



ChatGPT

The text from the image appears to say "е Мая", which likely corresponds to "of May" in English, considering the context and the style of the text, which suggests an older form of cursive.

