Layout Analysis Dataset with SegmOnto

Thibault Clérice, Juliette Janes, Hugo Scheithauer, Sarah Bénière, Laurent Romary, Benoît Sagot

DH2024: "Reinvention & Responsibility"

Layout Analysis Dataset with SegmOnto

Thibault Clérice[1], Juliette Janès[1], Hugo Scheithauer[1], Sarah Bénière[1], Laurent Romary[1], and Benoît Sagot[1]

[1]ALMAnaCH - Automatic Language Modelling and Analysis & Computational Humanities, Inria, Paris, France

November 2023

Automatic Text Recognition (ATR) has had a profound impact by generating extensive digital corpora from texts predating the digital era (Camps et al. 2023). While these technologies are undeniably valuable for tasks like human reading, PDF searching, plain-text keyword spotting, and, to a lesser extent, concordance building, the insufficient curation applied to most digitized texts (e.g., Archive.org, Project Gutenberg) presents a significant challenge when the correct sequential order of the text is compromised. Misidentifying footnotes, running titles, or marginal text as the main body of the text introduces noise into the character masses.

In computer vision, Document Layout Analysis (DLA) or Document Layout Segmentation (DLS) aims to extract semantics from distinct sections within documents, identifying common types like "text block" or "table." Recently, DLA has gained traction in digital humanities projects seeking to move beyond text mining towards document production (Scheithauer et al. 2021). In 2021, Gabay, Camps, and Pinche 2021 introduced the SegmOnto vocabulary, offering a controlled syntax to describe document features across various types (printed, manuscripts, incunabula). With categories such as `MainZone` for main text bodies and distinctions like `MarginTextZone` or `RunningTitleZone`, SegmOnto facilitates contamination-free bag-of-words generation, considering non-textual elements like stamps and digitization artifacts. While widely used in diverse projects, from medieval manuscripts (Pinche 2022) to printed classical texts commentaries (Najem-Meyer and Romanello 2022), SegmOnto remains relatively high-level and extendable. Its three-part syntax includes a controlled first level (e.g., `MainZone`) but allows flexibility in secondary (subtype) and third (numbering) levels, offering suggestions like `MainZone:Column#A`.

In this paper, we showcase an extensive implementation of the SegmOnto vocabulary for Layout Analysis, driven by a three-fold objective:

- Advocating for the adoption of a TEI-Based second level of SegmOnto, concentrating on segmenting sub-blocks of content (paragraphs, lists, etc.) when applied to text;
- Facilitating the generation of large bodies of digital texts that adhere to the sequential nature of the original content, despite its nonsequential representation on paper;
- Charting the course for more text-centric pipelines by recognizing that certain contents demand additional post-processing, such as bibliographic or lexicon entries, tables, etc.

This implementation is embodied in the "Layout Analysis Dataset with SegmOnto" (LADaS), a highly diverse dataset containing 3300 manually annotated and cross-corrected digitized page images sourced from 2000 distinct documents. The dataset encapsulates a broad chronological range, spanning from the 17th century to the 21st, and encompasses a wide array of genres, including poetry, academic papers (both humanities and STEM), novels, plays, epistolary works, and government and law reports.

This abstract begins by examining the current status of prominent datasets in the field of DLA. It then proceeds to discuss the guidelines, including our additions to the SegmOnto ontology. Following this, we present details about our dataset and its various components. Lastly, we conduct an evaluation to assess its effectiveness, demonstrated through its application in training an object detection model.

1 Current datasets

In the realm of Layout Analysis, datasets play a pivotal role, and recent developments have seen the establishment of benchmarking datasets for Computer Vision (CV) applications. Datasets like PubLayNet (Zhong, Tang, and Yepes 2019) and DocBank (Li et al. 2020) stand out, primarily centering around scientific papers sourced from ArXiv or PubMed. Newer datasets, such as DocLayNet (Pfitzmann et al. 2022) and M6Doc (Cheng et al. 2023), broaden their scope to include a more diverse range of sources. This last dataset appears most similar to LADaS, with akin labels, but differing in depth and focusing solely on modern documents, with a significant focus put on scientific papers due to their prevalence in LaTeX formats. This format facilitates preprocessing, enabling the extraction of valuable visual classification information (Pfitzmann et al. 2022).

Dataset	Documents Type	Pages	Labels
PubLayNet (Zhong, Tang, and Yepes 2019)	Medical papers	360000	5
DocBank (Li et al. 2020)	STEM papers	50000	12
DocLayNet (Pfitzmann et al. 2022)	Various Modern Documents	80863	11
Scibank (Grijalva et al. 2022)	STEM papers	74435	12
PrimA (Antonacopoulos et al. 2009)	Various Modern Documents	1240	10
M6Doc (Cheng et al. 2023)	Various Modern Documents	9080	74
SCUT CAB (Hiuyi et al. 2023)	Chinese Ancient Books	4000	27
American Stories (Dell et al. 2023)	Historic Press	2200	7
HJD (Shen, Zhang, and Dell 2020)	19th-20th Japanese Documents	2271	7
Gallicorpora (Sagot et al. 2022)	Literary books and manuscripts	981	15
HORAE (Boillet et al. 2019)	Books of hours	500	13
Ajax Multicommentary (Romanello, Sven, and Robertson 2021)	19th Critical Editions	300	18

Table 1: Comparison of Layout Analysis Datasets with Provided Page and Label Counts.

The diverse sources in layout analysis datasets often result in generic and numerous annotations using custom labels, with some datasets employing a generic approach, while others provide detailed zone descriptions, sometimes leading to more than to 70 labels (*cf.* Table 1). Materials commonly encountered in Digital Humanities (DH) endeavors, including historical press, monographs, manuscripts, and critical editions, have been addressed across different projects. In projects conducted after the introduction of SegmOnto, it has been utilized with a focus on standardization. However, both DH and CV projects typically lack the provision of a cross-genre, diachronic corpus aimed at text structuration.

2 Guidelines

While SegmOnto offers a set of zone types for distinguishing noise from the main body of text, it lacks specifications regarding the scope of annotation (e.g., whether the main zone should apply to the entire column or each paragraph individually) and lacks tools for normalized classification of sub-elements within the first level (e.g., paragraphs, lists). To construct our guidelines and class set, we selected subclasses based on the availability of a corresponding TEI class, the visual distinguishability of an element, and the relevance for post-processing information separation (e.g., `GraphiZone:Legend` and `GraphicZone`).

The most comprehensive main category in our subset is the `MainZone`, aligning with our primary textual focus. We distinguish groups of lines (‹lg›), paragraphs (‹p›), lists (‹item› within a ‹list›), and headings (‹title›). Although headings may resemble SegmOnto's `HeadingLine`, our application of zones to smaller text portions necessitates their representation. For many of these categories, as well as `MarginTextZone:Notes`, a third level marks the continuation of an element either

in a previous page or column, when this can be concluded from the layout (see Figure 1).

Finally, for the `TableZone` and `GraphicZone`, their sublevel `Legend` are instrumental in providing context for tabular and graphical elements. It enhances their interpretability. Furthermore, our annotations encompass specific subclasses such as `GraphicZone:P`, which focuses on annotating textual content within graphical elements. Notably, our design emphasizes the necessity for the main classes, such as `GraphicZone` and `TableZone`, to overlap with the legend subclasses. This strategic approach ensures a nuanced representation of the intricate relationships between describing textual content and graphical components, contributing to a more comprehensive understanding of document layout and structure in our dataset (see the complete set of classes in appendix, Table 5).

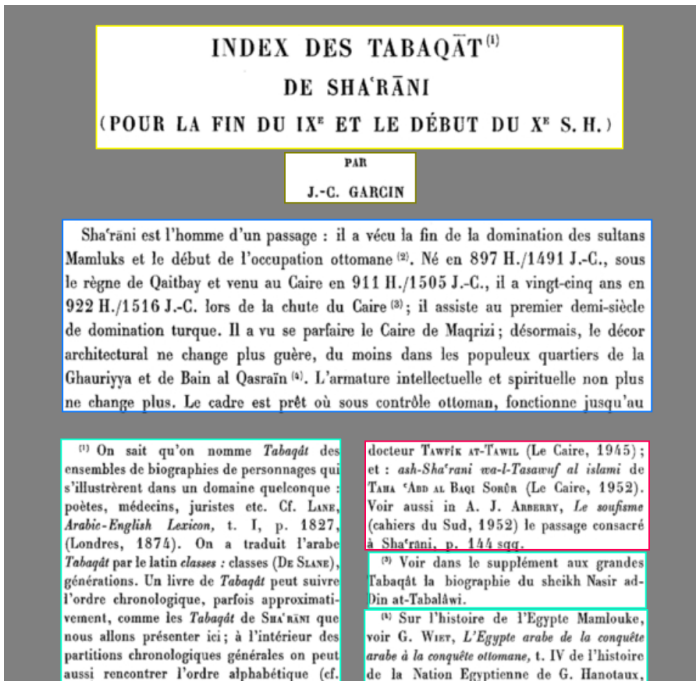


Figure 1: Example of Continued third level used, for a margin text note that continues on a second column. The lack of indentation marks the continuation here.

3 Content

Our dataset comprises five subsets (*cf.* Table 2, Figure 2 for its temporal coverage and Figure 4 for sample images), with annotations primarily carried out in an iterative

process on the Roboflow platform (Dwyer, Nelson, Solawetz, et al. 2022), involving annotation, training, and correction. Four distinct annotators contributed to the data annotations, and all annotations underwent proofreading by at least one additional annotator other than the original corrector. The object detection models were developed using the YOLOv8 architecture and associated software, subsequently integrated for in-app utilization on Roboflow. The final dataset is overly biased towards prose and paragraphs (see Table 5.)

The Gallica *Monographies* subset gathers 1785 books from the *Bibliothèque nationale de France* (BnF) and specifically from the OCR dump they provide on their data platform (*Gallica: bibliothèque numérique* n.d.). It contains various genres but it is mostly centered around genres containing paragraphs and has very few plays and poetry.

The "*DataCatalogue*" subset comprises 713 auction sales catalogs housed in the BnF and the Institut national d'histoire de l'art (INHA). These printed publications offer detailed information about auctioned items (numismatic, books, antiques, artworks, furniture, and luxury items), presenting structured entries followed by illustrations with distinctive layouts.

The *Persée* subset is derived from the French academic digital library and was curated by randomly selecting images from documents described on their OAI PMH server. Encompassing the period from 1824 to 2015, it includes diverse academic (STEM or HASS) papers and *miscellanea*. The layouts are very diverse, featuring unique topologies.

The *Fingers* subset has been built around the idea of adding noise in the dataset, we poorly digitized books using a book scanner but leaving fingers, background, and bent pages in the shot of the camera. These are meant to replicate more on-site issues and provide the ability for models to answer different needs, including the need of a researcher or student at a library to extract the text of a book.

Miscellanea are books or series of interest to the authors.

4 Evaluation

To assess the potential of our data, we trained an object detection model utilizing YOLOv8l with an 85/10/5% split and data augmentation on the training set (*cf.*

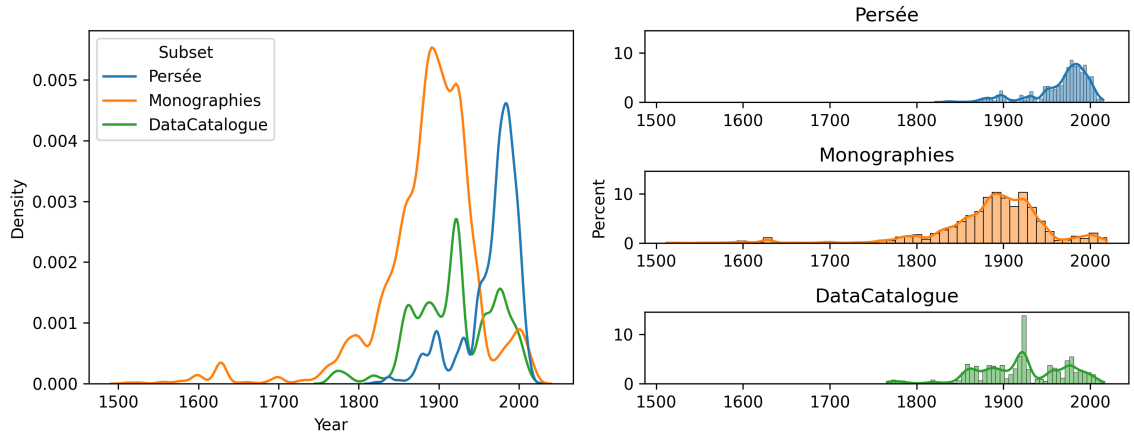


Figure 2: Corpus distribution over years for the main subsets. Subsets' graph uses the percentage of each subset

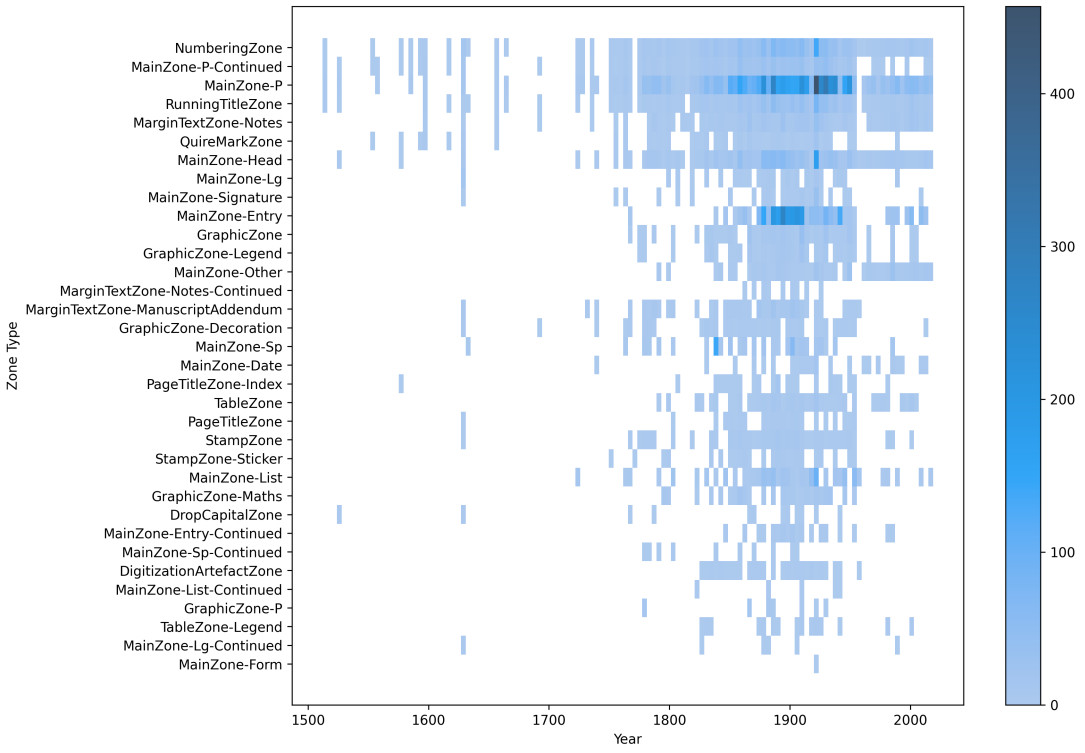
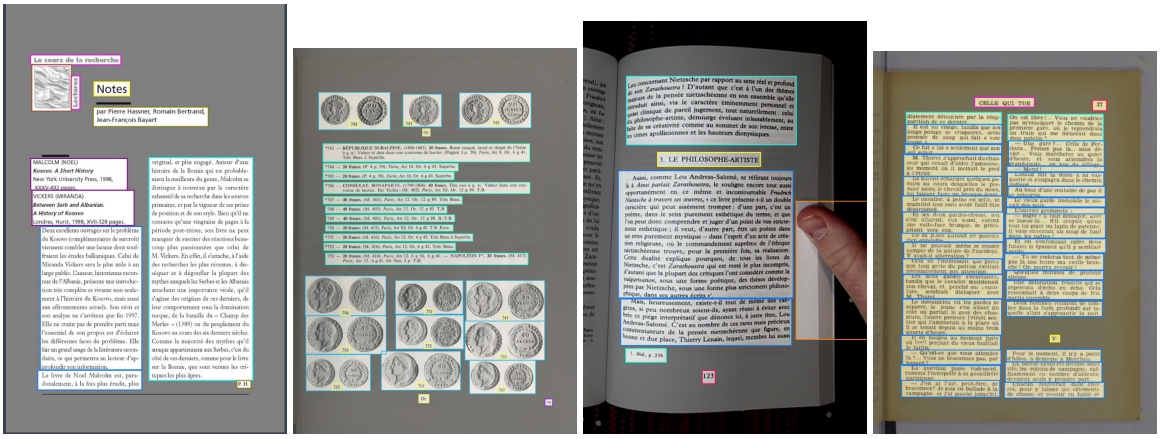


Figure 3: Presence of classes across time on the *Monographies* subset, which shows an important bias toward `MainZone:P` `RunningTitleZone`, and `MarginTextZone:Notes` with a constant presence in the corpus.

Subset	Genre	Annotated pages	Random	Max. samples per doc.	Century Start	Century End	Presence of other scripts than Latin
Monographies	Misc.	1991	✓	1	17	21	
DataCatalogue	Catalogs	750	✓	2	18	21	
Persée	Academic	815	✓	1	19	21	✓
Fingers	Misc.	60	✓	2	20	21	
Miscallanea	Misc.	8		1	19	19	✓

Table 2: LADaS Subsets properties.



(a) *Persée* (b) *DataCatalogue* (c) *Fingers* (d) *Monographies*

Figure 4: Example images from subsets.

appendix). The model[1] provides a strong score and shows an ability to detect most kinds of elements quite efficiently, except for rare items in the corpus, such as ‹sp›, ‹lg› and ‹list› (see Table 3). The third SegmOnto level which we use for element continuation (e.g. `MainZone:P#continued`) gets mostly captured, or at worst confused with the non-continuation equivalent (see Figure 5). The overall score (61.9 mAP50) paints a good picture and also provides promises for a bigger dataset in the future.

5 Conclusion

This paper presents two pivotal contributions to our community: (1) a set of SegmOnto guidelines that elevates text structuration beyond mere denoising of sequential data and (2) an initial set of images suitable for benchmarking architectures or training models for data extraction in extensive document collections. While a

[1]A demo of it used in a pipeline at `https://defi-colaf.github.io/TEIminatOCR/`.

Class	Images	Instances	Box(P)	Box(R)	mAP50	mAP50-95
all	288	2516	0.602	0.612	0.619	0.414
DigitizationArtefactZone	288	9	0.59	0.778	0.714	0.417
DropCapitalZone	288	7	0.749	0.852	0.798	0.443
GraphicZone	288	30	0.721	0.633	0.696	0.548
GraphicZone:Decoration	288	18	0.748	0.659	0.746	0.344
GraphicZone:Legend	288	25	0.694	0.4	0.461	0.274
GraphicZone:Maths	288	17	0.268	0.412	0.31	0.209
MainZone:Date	288	7	0.682	0.714	0.765	0.357
MainZone:Entry	288	393	0.874	0.888	0.909	0.695
MainZone:Entry#Continued	288	6	1	0.366	0.588	0.379
MainZone:Form	288	1	0	0	0	0
MainZone:Head	288	274	0.738	0.77	0.79	0.525
MainZone:Lg	288	6	0.619	1	0.901	0.728
MainZone:List	288	97	0.498	0.412	0.381	0.258
MainZone:Other	288	22	0.352	0.273	0.204	0.0952
MainZone:P	288	900	0.822	0.92	0.895	0.787
MainZone:P#Continued	288	98	0.744	0.847	0.842	0.773
MainZone:Signature	288	62	0.676	0.726	0.731	0.402
MainZone:Sp	288	26	0.713	0.192	0.339	0.241
MainZone:Sp#Continued	288	1	0	0	0	0
MarginTextZone:ManuscriptAddendum	288	10	0.75	0.899	0.855	0.419
MarginTextZone:Notes	288	73	0.533	0.616	0.608	0.424
MarginTextZone:Notes#Continued	288	2	0	0	0.108	0.097
NumberingZone	288	214	0.878	0.93	0.93	0.45
PageTitleZone	288	6	0.553	1	0.876	0.665
PageTitleZone:Index	288	11	0.681	0.818	0.787	0.677
QuireMarkZone	288	10	0.308	0.5	0.418	0.201
RunningTitleZone	288	141	0.858	0.904	0.888	0.532
StampZone	288	14	0.704	0.714	0.757	0.566
StampZone:Sticker	288	7	0.558	0.714	0.765	0.54
TableZone	288	23	0.77	0.783	0.698	0.523
TableZone:Legend	288	6	0.582	0.249	0.417	0.259

Table 3: Scores on the test set. We put in bold the categories that are both useful for text extraction and getting good scores. We can see that ‹sp›, ‹lg›, and ‹list› are yet to be well recognized but we expect these scores to be driven by the rarity of those classes in the training set.

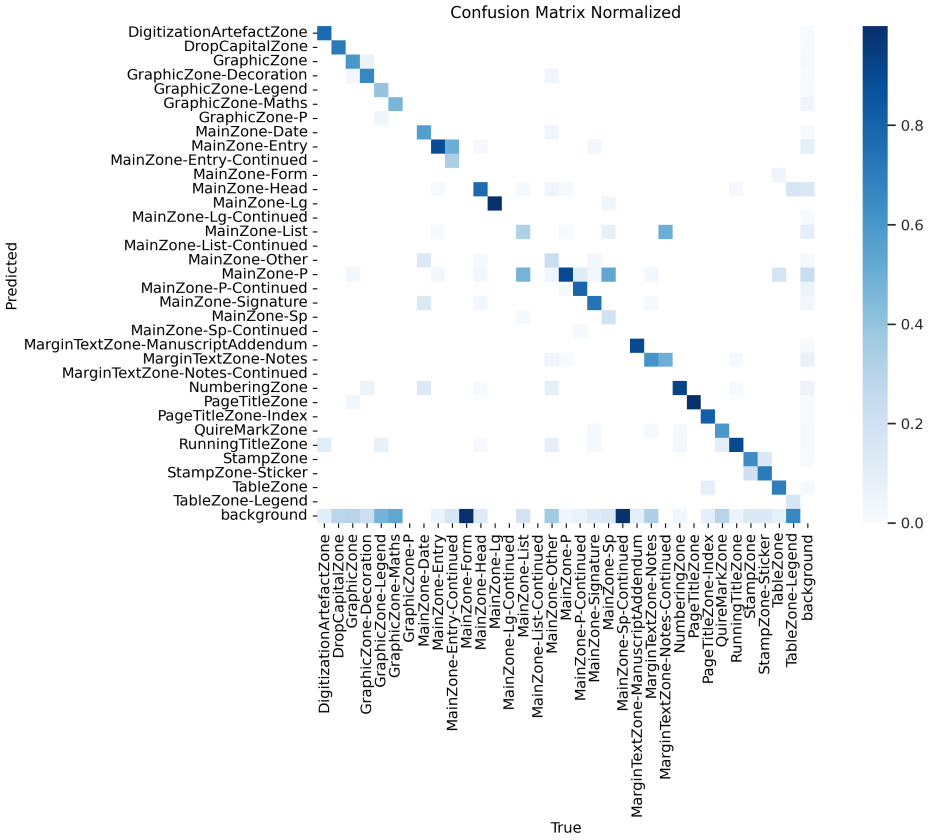


Figure 5: Confusion matrix, "background" means that the object was not detected. Most inter-classes confusion happens between items with a third level (e.g. `MainZone:Entry` and `MainZone:Entry#Continued`, as well as between main parts of the text: P, Other, List, and SP, with P being the class they are confused with.

comprehensive description of the guidelines is anticipated in the upcoming year, this paper offers a succinct preview of their potential and performance, underscored by a modest 3000 training samples.

However, our current project's focus is primarily on documents written in the Latin script and printed in Western Europe. While this limitation is acknowledged, we aim to overcome it through the refinement of guidelines, applying them to various printing traditions, and fostering collaboration with researchers specializing in diverse practices. This proactive approach ensures a more robust and inclusive framework for text structuration, paving the way for broader applications across different linguistic and printing contexts.

References

Antonacopoulos, Apostolos et al. (2009). "A Realistic Dataset for Performance Evaluation of Document Layout Analysis". en. In: *2009 10th International Conference on Document Analysis and Recognition*. Barcelona, Spain: IEEE, pp. 296–300. ISBN: 978-1-4244-4500-4. DOI: 10.1109/ICDAR.2009.271. URL: http://ieeexplore.ieee.org/document/5277696/ (visited on 11/14/2023).

Boillet, Mélodie et al. (Sept. 2019). "HORAE: an annotated dataset of books of hours". en. In: *Proceedings of the 5th International Workshop on Historical Document Imaging and Processing*. arXiv:2012.00351 [cs], pp. 7–12. DOI: 10.1145/3352631.3352633. URL: http://arxiv.org/abs/2012.00351 (visited on 11/27/2023).

Camps, Jean-Baptiste et al. (Dec. 2023). "Make Love or War? Monitoring the Thematic Evolution of Medieval French Narratives". In: *Computational Humanities Research (CHR 2023)*. CEUR Workshop Proceedings. URL: https://enc.hal.science/hal-04250657.

Cheng, Hiuyi et al. (June 2023). "M6Doc: A Large-Scale Multi-Format, Multi-Type, Multi-Layout, Multi-Language, Multi-Annotation Category Dataset for Modern Document Layout Analysis". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15138–15147. URL: https://github.com/HCIILAB/M6Doc.

Dell, Melissa et al. (Aug. 2023). *American Stories: A Large-Scale Structured Text Dataset of Historical U.S. Newspapers*. arXiv:2308.12477 [cs, econ, q-fin]. DOI: 10.48550/arXiv.2308.12477. URL: http://arxiv.org/abs/2308.12477 (visited on 11/17/2023).

Dwyer, Brad, Joseph Nelson, Jacob Solawetz, et al. (2022). *Roboflow (version 1.0) [software]*.

Gabay, Simon, Jean-Baptiste Camps, and Ariane Pinche (Nov. 2021). "SegmOnto". In: *Création de modèle(s) HTR pour les documents médiévaux en ancien français et moyen français entre le Xe-XIVe siècle*. Ecole nationale des chartes | PSL. Paris, France. URL: https://hal.science/hal-03481089.

Gallica: bibliothèque numérique (n.d.). Online. URL: https://gallica.bnf.fr.

Grijalva, Felipe et al. (2022). *SciBank: A Large Dataset of Annotated Scientific Paper Regions for Document Layout Analysis*. DOI: 10.21227/2yex-bt23. URL: https://dx.doi.org/10.21227/2yex-bt23.

Hiuyi, Cheng et al. (Nov. 2023). *SCUT CAB: A new Benchmark Dataset of Ancient Chinese Books with Complexe Layouts for Document Layout Analysis*. original-date: 2022-08-31T08:25:48Z. URL: https://github.com/HCIILAB/SCUT-CAB_Dataset_Release (visited on 11/23/2023).

Li, Minghao et al. (2020). *DocBank: A Benchmark Dataset for Document Layout Analysis*. arXiv: 2006.01038 [cs.CL]. URL: https://github.com/doc-analysis/DocBank.

Najem-Meyer, Sven and Matteo Romanello (2022). "Page Layout Analysis of Text-heavy Historical Documents: a Comparison of Textual and Visual Approaches". In: CEUR Workshop Proceedings. 3290, pp. 36–54. DOI: https://doi.org/10.48550/arXiv.2212.13924. URL: http://infoscience.epfl.ch/record/299775.

Pfitzmann, Birgit et al. (Aug. 2022). "DocLayNet: A Large Human-Annotated Dataset for Document-Layout Analysis". In: *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. arXiv:2206.01062 [cs], pp. 3743–3751. DOI: 10.1145/3534678.3539043. URL: http://arxiv.org/abs/2206.01062 (visited on 11/14/2023).

Pinche, Ariane (June 2022). *Cremma Medieval*. URL: https://github.com/HTR-United/cremma-medieval.

Romanello, Matteo, Najem-Meyer Sven, and Bruce Robertson (2021). "Optical Character Recognition of 19th Century Classical Commentaries: The Current State of Affairs". In: *The 6th International Workshop on Historical Document Imaging and Processing (HIP '21)*. Lausanne: Association for Computing Machinery. DOI: 10.1145/3476887.3476911.

Sagot, Benoît et al. (2022). "Galli(corpor)a: extraction, annotation et diffusion de l'information textuelle et visuelle en diachronie longue". In: URL: https://github.com/Gallicorpora/.

Scheithauer, Hugo et al. (Oct. 2021). "From page to content – which TEI representation for HTR output?" In: *Next Gen TEI, 2021 - TEI Conference and Members' Meeting*. Weaton (virtual), United States. URL: https://hal.science/hal-03380807.

Shen, Zejiang, Kaixuan Zhang, and Melissa Dell (2020). "A Large Dataset of Historical Japanese Documents With Complex Layouts". In: pp. 548–549. URL: https://openaccess.thecvf.com/content_CVPRW_2020/html/w34/Shen_A_Large_Dataset_of_Historical_Japanese_Documents_With_Complex_Layouts_CVPRW_2020_paper.html (visited on 11/23/2023).

Zhong, Xu, Jianbin Tang, and Antonio Jimeno Yepes (Sept. 2019). "PubLayNet: largest dataset ever for document layout analysis". In: *2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, pp. 1015–1022. DOI: 10.1109/ICDAR.2019.00166. URL: https://github.com/ibm-aur-nlp/PubLayNet.

Dataset	Type of documents	Digitization	Pages	Human annotators	Classes Number	Examples	Reusable
Scibank	STEM papers	No	74435	No	12	abstract, textblocks, caption, equation, inline equation, figure, keywords, reference, section, subsection, table, title.	CC-BY
PubLayNet	Medical papers	No	360000	No	5	Text, Title, List, Table, Figure	CDLA - Permis 1.0
M6Doc	Textbook, test paper, magazine, newspaper, scientific papers, notes, book	Mixed	9080	Yes	74	Head, SubHead, Headline, Title, Paragraph, Section, Section Title, Chapter Title, Footer, Footnote, Ordered List, Nonordered List, Table, Caption...	CC BY-NC-ND 4.0
PrimA	Magazine, technical journals, forms, bank statements, ads	Mixed	1240	Yes	10	Text, image, line drawing, graphic, table, chart, separator, maths, noise, frame	None
DocLayNet	financial reports, scientific articles, laws and regulations, government tenders, manuals, patents	Mixed	80863	Yes	11	Caption, Footnote, Formula, ListItem, Page Footer, Page Header, Picture, Section Header, Table, Text, Title	CDLA - Permis 1.0
DocBank	STEM papers	No	500000	No	12	Abstract, Author, Caption, Equation, Figure, Footer, List, Paragraph, Reference, Section, Table, Title	None
SCUT CAB	Chinese Ancient Books	Yes	4000	Yes	27	bibliography, header, pagebox, book number, figure, author, title, ear note, chapter title...	CC-BY-NC-ND 4.0
Historial Japanese Dataset	Japanese Historical Doc (19th-20th)	Yes	2271	Yes	7	page frame, row, title region, text region, title, subtitle, other	APACHE 2.0
Gallicorpora	Literary books & manuscripts	Yes	981	Yes	15	SegmOnto (level 1)	
American Stories	Historic Press	Yes	2200	Yes	7	Author, Article, Header, Headline, Table, Masthead, Cartoon/ads	CC-BY
HORAE	Books of hours	Yes	500	Yes	13	Page, textregion, bordertext, textline, miniature, decorated border, illustrated border, initiale (simple, decorated, historiated), line filer, music notation, ornamentation	CC-BY
AJAX Multicommentary Dataset	Critical editions 19th	Yes	300	Yes	18	Commentary, critical apparatus, footnotes, page number, text number, bibliography, handwritten marginalia, index, others, printed marginalia, table of contents, title, translation, appendix, introduction, preface, primary text, running header	CC-BY

Table 4: Detailed Comparison of Layout Analysis Datasets (Table 1)

Training parameters

- 100 epochs
- Image Size: 960 pixels
- **Image augmentation**
  - Outputs per training example: 3
  - Rotation: Between -3° and +3°
  - Shear: ±3° Horizontal, ±3° Vertical
  - Exposure: Between -30% and +30%
  - Blur: Up to 1.5px
  - Noise: Up to 2% of pixels

Label	Description	Cont.
DigitalArtefactZone	Element added during the document's digitization	
DropCapitalZone	Initial capital at the beginning of a text	
GraphicZone:Decoration	Decorative image without text (banner, ornamental dividers, etc.)	
GraphicZone	Illustration such as images, photographies, or graphs, and their legends	
GraphicZone:Legend	Textual caption for a GraphicZone or a GraphicZone:Maths	
GraphicZone:Maths	Mathematical formula not in a paragraph	
GraphicZone:P	Paragraph in a GraphicZone	
MainZone:Date	Date isolated from the rest of a paragraph, at the beginning or the end	
MainZone:Entry	Structured text such as bibliographic and catalogs' entries	✓
MainZone:Form	Form to complete	
MainZone:Head	Title of a part or a chapter on a non-title page	
MainZone:List	Item of an enumerated or ordered list	✓
MainZone:Lg	Group of verse lines	✓
MainZone:Other	When none of the previous labels can be used	
MainZone:P	Paragraph	✓
MainZone:Signature	Name of a person visually separated from the rest of the text	
MainZone:Sp	Performance text, including stage directions and name of the person speaking in case of a theatrical text	✓
MarginTextZone:ManuscriptAddendum	Handwritten note	
MarginTextZone:Notes	Footnote and margin note	✓
NumberingZone	Page Numbering	
PageTitleZone	Front Cover of the document	
PageTitleZone:Index	Table of contents	
QuiremarkZone	Information at the bottom of a page which is not page number or running title	
RunningTitleZone	Title or abbreviated title at the top or the bottom of the page	
StampZone	Library Stamp	
StampZone:Sticker	Sticker on library books	
TableZone	Table and its legend	
TableZone:Legend	Textual caption for a TableZone	

Table 5: LADaS annotation schema. Cont. indicates that the class can have a third SegmOnto level indicating that the element extends from a preceding zone, *e.g.* `MainZone:P#Continued`.