International Journal on Document Analysis and Recognition (IJDAR)
https://doi.org/10.1007/s10032-026-00585-7

ORIGINAL PAPER

U-DIADS-TL: a novel dataset for text line segmentation in historical manuscripts

Silvia Zottin[1] · Axel De Nardin[1] · Claudio Piciarelli[1] · Gian Luca Foresti[1]

Received: 6 June 2025 / Revised: 16 March 2026 / Accepted: 4 April 2026

Abstract

Text line segmentation in historical documents remains a significant challenge due to degraded manuscripts, complex layouts, and diverse handwriting styles. Developing robust computational methods is hindered by the scarcity of high-quality ground truth annotations, which require expert knowledge and are time-intensive to produce. Few-shot learning has emerged as a promising solution by enabling model training with minimal annotated data, yet its application to historical document analysis is still largely unexplored. To address this limitation, we introduce U-DIADS-TL (Uniud - Document Image Analysis DataSet - Text Line), a dataset specifically designed for text line segmentation in ancient manuscripts. U-DIADS-TL provides noise-free annotations with non-overlapping text elements and accommodates diverse document structures, including multi-column layouts. To encourage few-shot learning approaches, we offer only three training images, allowing researchers to develop segmentation models that can generalize from limited supervision. Our dataset serves as a critical bridge between deep learning and historical document analysis, fostering the creation of efficient, adaptable segmentation models for real-world applications.

Keywords Text line segmentation · Text line dataset · Document layout analysis · Few-shot analysis

1 Introduction

The analysis of historical document images presents unique challenges due to various types of degradation, including faded ink, holes, stains, and bleed-through. These issues, combined with complex layouts and diverse handwriting styles, make the processing of such manuscripts difficult for both humanities scholars and computational methods. As a result, the design and development of specialized techniques are essential for effectively extracting and analyzing textual and structural elements in historical documents. A fundamental step in this process is represented by the layout analysis of the documents, which aims to determine their physical and logical structure. One key aspect of layout analysis is text line segmentation, which focuses on identifying and localizing individual lines of text within a document image. This step is crucial for subsequent tasks such as optical character recognition and handwritten text recognition, making it a critical component of historical document analysis.

One of the major challenges in developing robust document analysis methods is the need for high-quality ground truth annotations. Creating noise-free ground truth requires expert knowledge and significant time and resources, limiting the availability of annotated datasets. To address this issue, few-shot learning has emerged as a viable solution, enabling models to learn effectively from a very limited number of training samples. Few-shot learning is particularly beneficial for historical document analysis, where obtaining extensive labeled datasets is impractical. Despite the importance of layout analysis, few works in the literature have tackled this task by relying on approaches specifically designed for few-shot learning or alternative methods that mitigate the challenges of acquiring precise annotations. Some recent works have explored few-shot learning approaches [1–6], as well as

✉ Silvia Zottin
silvia.zottin@uniud.it

Axel De Nardin
axel.denardin@uniud.it

Claudio Piciarelli
claudio.piciarelli@uniud.it

Gian Luca Foresti
gianluca.foresti@uniud.it

[1] AVML Lab, Department of Mathematics, Computer Science and Physics, University of Udine, Via delle Scienze 206, 33100 Udine, Italy

Published online: 18 April 2026

transfer learning techniques [7–9] and unsupervised methods to reduce the reliance on labeled data [10–12]. However, the research on few-shot learning approaches specifically tailored for historical document layout analysis remains limited.

As a way to partially fill this gap, in this work we introduce the Uniud - Document Image Analysis DataSet - Text Line version (U-DIADS-TL), a dataset specifically designed for text line segmentation in ancient manuscripts. Unlike many existing text line segmentation datasets, U-DIADS-TL is characterized by a high degree of precision in the text line annotations, ensuring high accuracy and reliability. It features non-overlapping elements that provide clean and unambiguous segmentation, along with noise-free annotations that enhance the usability of the dataset. Furthermore, the dataset includes heterogeneously oriented text lines, accommodating diverse layouts and structures, and multi-column page structures that reflect the complexity of ancient documents.

To promote few-shot learning and to address the challenges of ground truth generation, we provide only three images as the training set for each document class, making our approach highly relevant for real-world applications. This setup ensures that models trained on U-DIADS-TL can be effectively used to assist humanities scholars in the analysis of historical manuscripts.

The main contributions of this work can be summarized as follows:

- We propose U-DIADS-TL, a novel dataset for text line segmentation in historical manuscripts, providing high-precision, noise-free annotations and capturing challenges such as faded ink, stains, bleed-through, diverse handwriting styles, multi-ink text lines, varying interline spacing, and marginal comments.
- We introduce a few-shot learning setup for the text line segmentation task, with only three training images per manuscript, enabling the evaluation of segmentation models in low-resource scenarios.
- We provide baseline results using established segmentation models as well as state-of-the-art approach to serve as a reference for future research and to highlight the challenges posed by the dataset.

The remainder of this paper is organized as follows: Sect. 2 provides an overview of historical handwritten datasets available in the literature for the text line segmentation task. A comprehensive description of the U-DIADS-TL dataset is presented in Sect. 3. Section 4 outlines the benchmark results of the dataset. Finally, Sect. 5 concludes the paper with key insights and directions for future work.

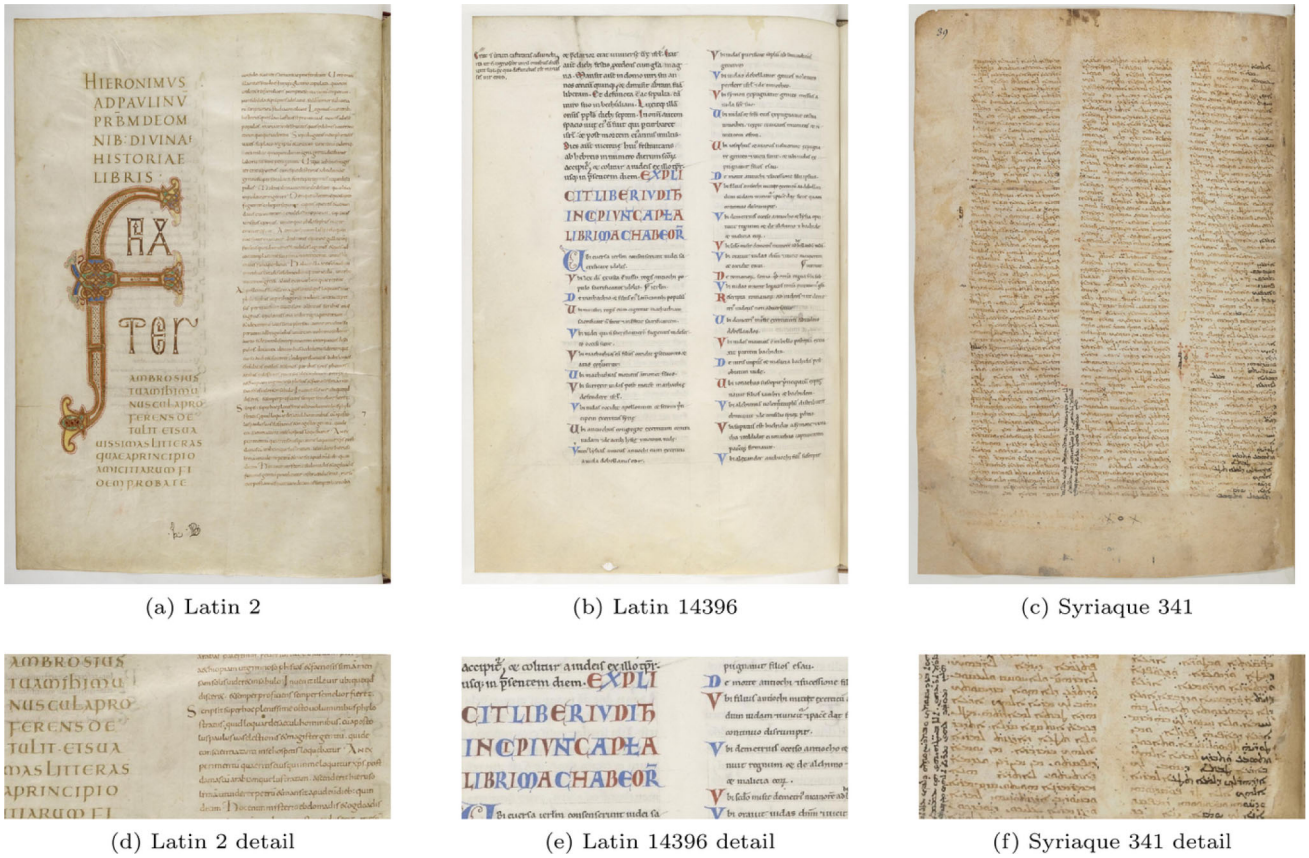
2 Related work

Several datasets have been developed in the literature to address the text line segmentation task in ancient manuscripts. A systematic literature review of existing datasets for historical document image analysis is presented in [13].

DIVA-HisDB [14] represents the most prominent example of dataset crafted for this purpose, and consists of 150 annotated images from three medieval manuscripts: CSG18, CSG863, and CB55, with 50 pages per manuscript. It is divided into training, validation, and test sets, containing 30, 10, and 10 pages per manuscript, respectively. This dataset supports tasks such as layout segmentation, baseline detection, and text line segmentation. Another notable dataset is presented in RASM2018 [15]. The dataset was introduced for the RASM 2018 competition, which focused on recognizing Arabic historical scientific manuscripts. It includes 15 single-column pages for training and 85 for evaluation, supporting tasks such as page segmentation, text line detection, and OCR.

The Pinkas dataset [16] was created from a historical Hebrew manuscript documenting Jewish communities in Europe between 1500 and 1800. It comprises 30 digitized pages, with ground truth annotations available in PAGE format. A larger dataset is represented by Digital Peter [17], which contains 662 pages of manuscripts written by Peter the Great between 1709 and 1713. The dataset is divided into 595 pages for training and 67 for testing, all with annotations. For Chinese historical documents, the ICDAR 2019 HDRC-CHINESE dataset [18] was introduced as part of the Historical Document Reading Challenge on Large Chinese Structured Family Records. It includes approximately 10,000 historical Chinese family record pages and supports tasks such as text recognition on extracted lines, pixel-level layout analysis, and text line detection and recognition. The PHTD dataset [19] focuses on Persian handwritten text, containing 140 handwritten documents across three categories, written by 40 individuals. It includes 1,787 text lines and 27,073 words/subwords. Many of the documents feature overlapping or touching text lines, with an average of 13 lines per page.

Another relevant dataset is GRPOLY-DB [20] (Greek Polytonic Database), which consists of images of printed and handwritten documents using the old polytonic Greek script from 1838 to 1977. It is divided into four subsets: GRPOLY-DB-Handwritten, GRPOLY-DB-MachinePrinted-A, GRPOLY-DB-MachinePrinted-B, and GRPOLY-DB-MachinePrinted-C. The BH2M dataset [21] (Barcelona Historical Handwritten Marriages Database) provides 174 handwritten marriage record pages written in Old Catalan between 1617 and 1619. The dataset is divided into 100 pages for training, 34 for validation, and 40 for testing, with ground truth annotations available for layout analysis, text transcrip-



(a) Latin 2 (b) Latin 14396 (c) Syriaque 341

(d) Latin 2 detail (e) Latin 14396 detail (f) Syriaque 341 detail

Fig. 1 Samples from the 3 manuscripts (Latin 2, Latin 14396, and Syriaque 341) present in U-DIADS-TL. Figure (**a**) - (**c**) show a full page for each manuscripts, while Fig. (**d**) - (**f**) show a detail extracted from each of them

tion, and semantic analysis. For manuscripts with complex layouts, the VML-MOC dataset [22] (Visual Media Lab - Multiply Oriented and Curved) includes 30 pages from historical manuscripts. Some manuscripts originate from a private library in Jerusalem, while others come from the Islamic manuscript digitization project at Leipzig University Library. This dataset contains handwritten documents with multiply oriented and curved text lines. Similarly, the VML-AHTE dataset [23] provides a benchmark for handwritten text lines with crowded diacritics, touching, and overlapping characters. It includes 20 training pages and 10 test pages, with full line-level annotations by native Arabic speakers.

In this paper, we introduce U-DIADS-TL, a novel dataset for text line segmentation of ancient manuscripts. This dataset is characterized by noiseless ground truth annotations and is specifically designed to foster further research on few-shot learning approaches. To our knowledge, only one other dataset in the literature is structured in a few-shot setting, but with a primary focus on document layout segmentation [24]. Our goal is to promote advancements in this field while addressing challenges related to time constraints, annotation precision, and the involvement of domain experts for the creation and labeling of ground truth data that are essential for machine and deep learning approaches. A preliminary version of this dataset has been used in the ICDAR 2025 FEST Competition on FEw-Shot Text line segmentation of ancient handwritten documents [25].

3 Overview of U-DIADS-TL dataset

The U-DIADS-TL dataset was created through a collaborative effort between computer scientists and humanities scholars. It comprises three distinct ancient manuscripts (Latin 2, Latin 14396, and Syriaque 341), with 28 images selected from each. These unique color page images were carefully selected from each manuscript and divided into three subsets: 3 images for training, 10 for validation, and 15 for testing. The dataset is freely available for download on the U-DIADS-TL repository[1].

[1] https://sites.google.com/view/avml-lab-udiadstl.

3.1 Descriptions of the manuscripts

The images were sourced from the French digital library Gallica[2], and the manuscripts represent Latin and Syriac Bibles dating from the 6th to the 12th centuries A.D. A sample page from each manuscript, along with a corresponding detailed view, is presented in Fig. 1. Below is a detailed description of each manuscript:

- **Latin 2**[3]: Known as the *Second Bible of Charles the Bald*, this manuscript was created between A.D. 871 and 877 at the Abbey of Saint-Amand (Haute-France). It comprises 444 parchment pages, structured in a two-column layout, and features richly decorated illuminations and intricate scripts characteristic of Carolingian art.
- **Latin 14396**[4]: Often referred to as *Genesis-Kings, the First Volume*, this manuscript was produced between A.D. 1145 and 1150 at the Abbey of Saint-Victor (Paris). It contains 176 parchment pages, also organized in a two-column layout, with notable features, including elaborate initials and a distinct Romanesque script style.
- **Syriaque 341**[5]: Created between the 6th and 7th centuries A.D., this manuscript is thought to have originated from the Monastery of Baquqa in Iraq. It consists of 256 pages made of both parchment and paper, arranged in a unique three-column layout.

3.2 Dataset challenges

The U-DIADS-TL dataset presents several challenges that make it a valuable benchmark for text line segmentation in historical document analysis. One of the primary difficulties arises from the inherent complexity of ancient manuscripts, which exhibit diverse handwriting styles and significant differences in letter sizes and shapes. The presence of multi-oriented text lines and interline comments further complicates segmentation tasks, as traditional methods often struggle to adapt to non-horizontally oriented layouts.

Additionally, the manuscripts in U-DIADS-TL feature multi-column structures, requiring robust algorithms capable of distinguishing between closely spaced text regions. Moreover, the dataset spans multiple languages, each with distinct script characteristics. The segmentation process is further hindered by various forms of document degradation, such as faded ink, stains, holes, and bleed-through, which obscure text boundaries and introduce noise.

Fig. 2 Examples of difficulties and challenges present in document images from the U-DIADS-TL dataset: faded ink and stains (**a**), bleed-through (**b**), multi-oriented text lines (**c**), multi-ink colored text lines (**d**), diverse handwriting styles and size (**e**), and interlines and marginal comments (**f**)

Some visual examples of the difficulties and challenges present in document images in the U-DIADS-TL dataset are shown in Fig. 2.

Another key challenge is the limited availability of annotated training data. Since historical documents often require expert validation, large-scale manual annotation is infeasible. This constraint highlights the importance of few-shot learning approaches, where models must generalize effectively from a small number of labeled examples. Unlike existing text line segmentation datasets [14, 18, 21–23], U-DIADS-TL is designed to support the development and evaluation of segmentation models in low-data scenarios.

These manuscripts were carefully selected in collaboration with humanities scholars to prioritize their diverse and complex layouts. Each manuscript exhibits unique and dis-

[2] https://gallica.bnf.fr/

[3] https://gallica.bnf.fr/ark:/12148/btv1b8452767n.

[4] https://gallica.bnf.fr/ark:/12148/btv1b84429190.

[5] https://gallica.bnf.fr/ark:/12148/btv1b10527102b.

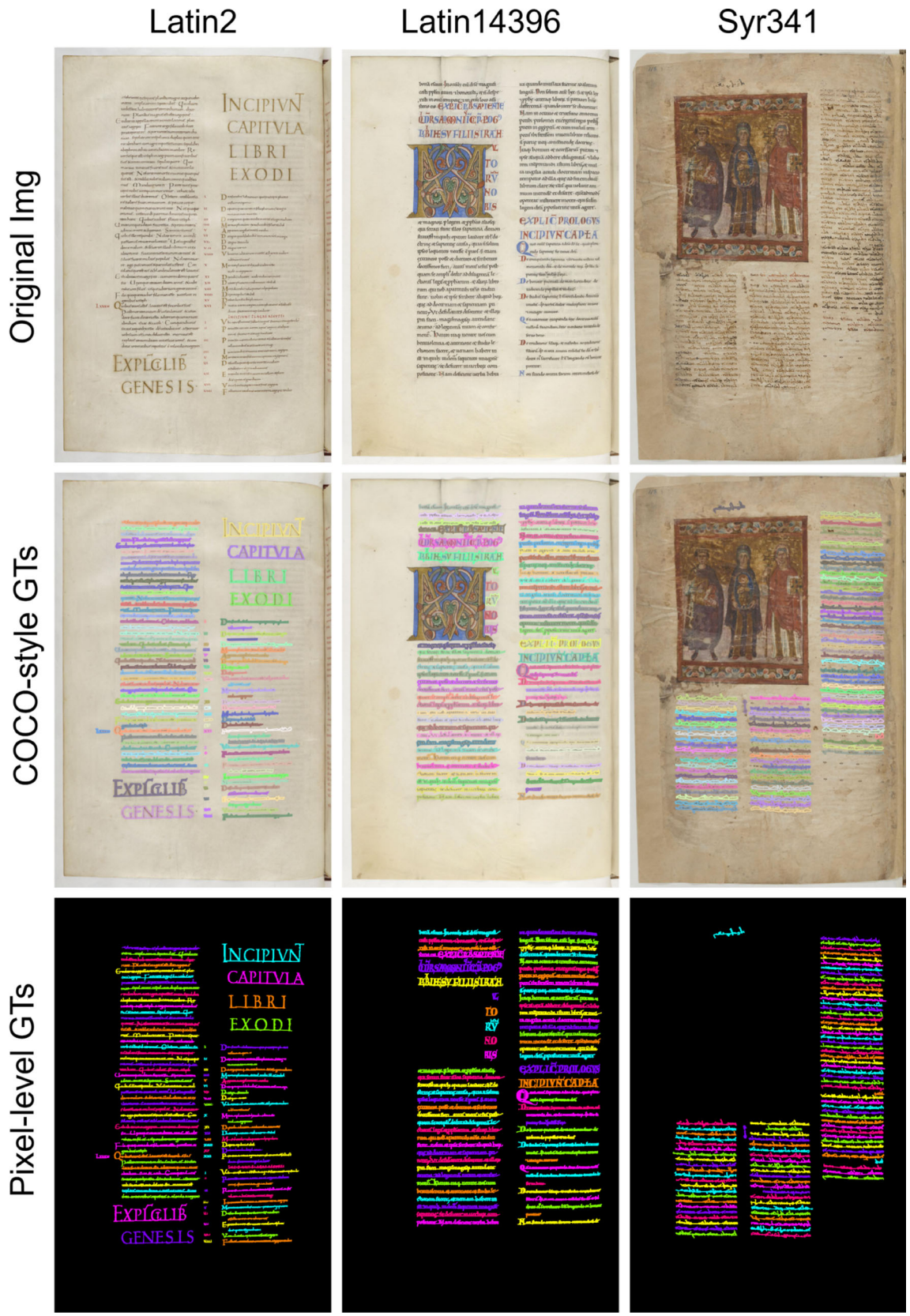


Fig. 3 Samples images (top row), COCO style GTs (middle row) and pixel-level GTs (bottom row) from the three manuscripts (Latin 2, Latin 14396, and Syriaque 341) included in U-DIADS-TL are presented. In the GT samples, each color represents a different text line segmentation instance

tinct graphical features that add depth to the dataset. Notably, Syriaque 341 stands out as the most challenging among the three. Its pages feature vertical comments, ink in varying colors, and a high degree of degradation due to aging and poor preservation conditions. These exceptional characteristics not only test the robustness of text line segmentation methods but also make the dataset particularly intriguing and valuable for advancing research in historical document analysis.

The combination of ancient, multi-oriented, multilingual, handwritten, and degraded documents, along with its design to support the development of few-shot approaches, makes U-DIADS-TL a particularly challenging benchmark. It pushes the boundaries of current segmentation methodologies and encourages the development of more adaptive and generalizable models.

3.3 Ground truth

Accurate Ground Truth (GT) information is essential for the automatic evaluation of any segmentation or recognition system. However, manual annotation is both time-consuming and error-prone, particularly in document segmentation, where intricate layout components must be carefully distinguished. To address this challenge, the annotation of the U-DIADS-TL dataset was carried out through a collaborative effort between humanities scholars and computer scientists. Each page is accompanied by detailed GT annotations, which distinguish between two classes: background and text lines.

We implemented a segmentation pipeline that alternates between humanities scholar annotation and algorithmic processing to optimize results. Initially, one image per manuscript was selected and binarized using the Sauvola thresholding technique [26] and morphological operators to provide humanities scholars with a starting point. Manuscript specialists then manually segmented these images. Using this manually annotated subset, we trained a machine learning model to generate a coarse segmentation of the entire dataset. This process followed the framework proposed in [27], which presents a high-performance one-shot pixel-precise document layout segmentation method. However, unlike [27], our segmentation focused on two distinct classes: text and background.

Finally, humanities scholar meticulously refined the text line annotations, comparing the generated masks with the original images to ensure accuracy. Despite leveraging computational assistance, the final annotations were always validated and corrected by humanities scholar, mitigating potential biases and errors in the dataset.

An important design decision in defining the GT for text lines was the introduction of a connecting line that ensures each text line forms a distinct connected component. This structural choice increases the versatility of the dataset, enabling the use of a broader range of segmentation models. In particular, it facilitates not only instance segmentation approaches, where each text line is treated as a separate instance, but also binary segmentation models. The latter do not require unique identifiers for individual lines; instead, the segmentation can rely on connected components to distinguish between different text lines, thereby simplifying evaluation and enhancing compatibility with various methods. The GTs are provided both as polygon-based annotations in COCO format and at the pixel level.

Table 1 Total amount of text lines for each manuscripts of U-DIADS-TL

#Number	Train	Validation	Test	Total
Latin 2	317	955	1469	2741
Latin 14396	232	785	1166	2183
Syriaque 341	463	1747	2626	4836

Figure 3 presents examples of the defined GT alongside their respective original images for each manuscript. In the GT images, each color represents a distinct text line segmentation instance.

The total number of text lines for each manuscript is summarized in Table 1.

The division of images into training, validation, and test splits is already defined in the released dataset, and any changes to this division would prevent direct comparison with the reported results.

4 Benchmark setup

To analyze our data and establish a benchmark for future studies, we evaluate the U-DIADS-TL dataset using three widely adopted deep learning-based segmentation methods: Fully Convolutional Network (FCN) [28], Pyramid Scene Parsing Network (PSPNet) [29], and DeepLabv3+ [30]. Additionally, we include the results obtained by a state-of-the-art supervised learning approach for text line segmentation presented by Barakat et al. [31]. To the best of our knowledge, no existing works have explored few-shot learning approaches for text line segmentation, and thus they are not included in our comparison.

The FCN is a pioneering deep learning architecture for semantic segmentation that replaces fully connected layers with fully convolutional layers, enabling dense predictions on images of arbitrary size. It follows an encoder-decoder structure, where the encoder extracts hierarchical features, and the decoder restores spatial resolution using upsampling layers and learned deconvolution filters. Skip connections enhance segmentation accuracy by combining coarse semantic infor-

Table 2 Performance comparison of different models on the U-DIADS-TL dataset (Line IU, Pixel IU, Detection Rate, Recognition Accuracy, and F-Measure)

Model	Metric↑	Latin 2	Latin 14396	Syriaque 341	Average
FCN [28]	LIU	0.449	0.493	0.156	0.366
	PIU	0.513	0.551	0.183	0.416
	DR	0.292	0.437	0.099	0.276
	RA	0.166	0.269	0.092	0.176
	FM	0.210	0.329	0.095	0.211
PSPNet [29]	LIU	0.402	0.519	0.020	0.314
	PIU	0.515	0.503	0.079	0.365
	DR	0.154	0.338	0.003	0.165
	RA	0.110	0.302	0.007	0.140
	FM	0.126	0.316	0.004	0.149
DeepLabv3+ [30]	LIU	0.558	0.582	0.230	**0.457**
	PIU	0.554	0.573	0.342	**0.490**
	DR	0.457	0.512	0.180	**0.383**
	RA	0.250	0.331	0.116	0.232
	FM	0.321	0.397	0.140	0.286
Barakat et al. [31]	LIU	0.581	0.578	0.029	0.396
	PIU	0.535	0.521	0.071	0.376
	DR	0.535	0.568	0.021	0.375
	RA	0.444	0.440	0.051	**0.312**
	FM	0.448	0.489	0.029	**0.322**

The best-performing results for the full dataset are shown in bold

mation from deep layers with fine spatial details from shallow layers, improving localization and boundary precision.

PSPNet is a semantic segmentation model designed to capture both local and global contextual information using a Pyramid Pooling Module. Built on a ResNet backbone, PSPNet extracts hierarchical features and enhances them by pooling information from multiple spatial scales. The decoder then refines these enriched feature maps through convolutional layers to produce high-resolution pixel-wise predictions.

DeepLabv3+ is a state-of-the-art semantic segmentation model that improves contextual understanding and spatial detail refinement through an encoder-decoder structure. Its encoder leverages atrous convolutions and the Atrous Spatial Pyramid Pooling module to capture multi-scale features efficiently. The decoder refines segmentation by upsampling encoder outputs and integrating lower-level feature maps, enhancing boundary precision and overall accuracy.

For these three segmentation models, those provided by the Segmentation Models PyTorch library [32] were used. We trained all networks from scratch using the same hyper-parameter setup and full-resolution images. All these models are trained under the same conditions, utilizing the Adam optimizer with a learning rate of 10^{-3} and a weight decay of 10^{-5}. Training runs for a maximum of 100 epochs, with an early stopping mechanism that monitors validation loss at each epoch and halts training if no improvement is observed for 20 consecutive epochs. Each model employs a residual network as its backbone, specifically the ResNet50 variant. All models were trained separately for each manuscript.

Finally, the approach proposed by Barakat et al. employs an FCN model, incorporating both pre-processing and post-processing stages. Specifically, a total of 50,000 and 6,000 random patches of size 320 × 320 were generated for the training and validation sets of each fold, respectively. Post-processing involves computing component orientations, segmenting the image into layers, applying directional morphological operations, and merging the results using a pixel-wise OR operation. This approach was re-implemented for our experiments while preserving the original setup described by the authors as closely as possible.

We performed all the presented experiments on an NVIDIA A100 GPU.

4.1 Evaluation protocol

To evaluate the selected models on our dataset as a means to provide a set of baseline results to build upon for future research, we selected five popular text line segmentation metrics commonly adopted in the literature [14, 33], namely Line Intersection over Union (LIU), Pixel over Union (PIU), Detection Rate (DR), Recognition Accuracy (RA), and F-Measure (FM).

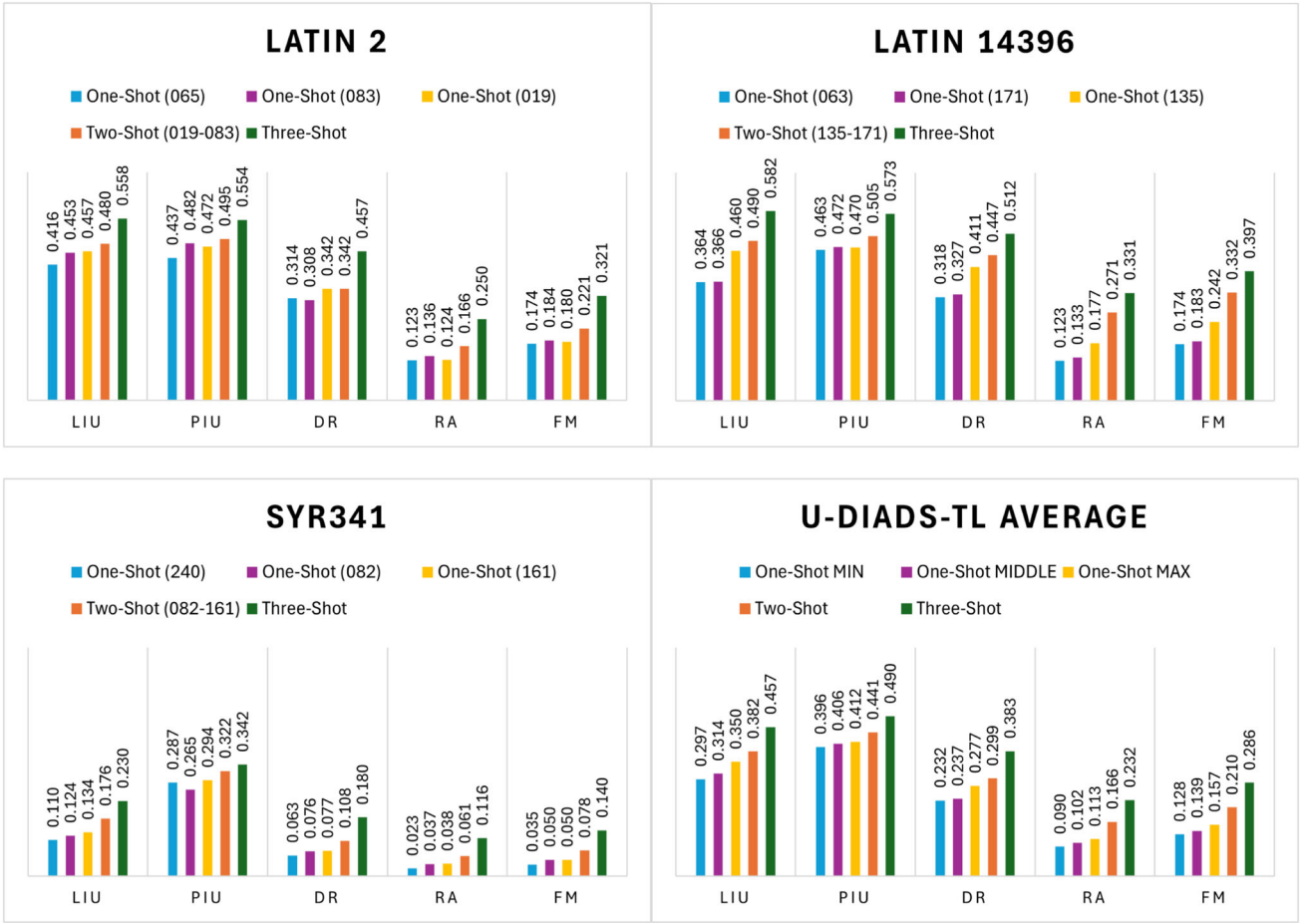


Fig. 4 Histograms illustrating performance metrics for one-, two-, and three-shot analysis across individual manuscripts and the overall U-DIADS-TL dataset with DeepLabv3+ model. The numbers inside the parenthesis represent the id of the corresponding instances of the dataset used during the training process

Line Intersection over Union and Pixel over Union are presented in the DIVA evaluation protocol [14] and is based on the Intersection over Union (IU) metric, defined as:

$$\mathrm{IU} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{1}$$

where TP, FP, and FN denote True Positives, False Positives, and False Negatives respectively.

Line IU evaluates IU at the line level, measuring how accurately entire lines are detected. TP corresponds to correctly detected lines, FP to extra (false positive) detected lines, and FN to missed (false negative) lines. A minimum overlap threshold of 75% is applied to determine matches between predicted and ground-truth text lines. Two text lines are considered a match if both pixel precision and recall exceed this threshold; otherwise, they are classified as FP (precision $<$ threshold) or FN (recall $<$ threshold).

A second set of metrics, namely Detection Rate (DR), Recognition Accuracy (RA), and F-Measure (FM) has been selected for further validation of the different approaches' performance on our benchmark. These evaluation criteria have been proposed by the ICFHR 2010 Handwriting Segmentation Contest [33]. These metrics rely on the MatchScore metric. Given an image, let R_i denote the points within the i-th detected line segment, G_j the points within the j-th ground truth line segment, and $T(p)$ the number of points in a set p. The MatchScore between a detected and ground truth segment is computed as:

$$\mathrm{MatchScore}(i, j) = \frac{T(G_j \cap R_i)}{T(G_j \cup R_i)} \tag{2}$$

A region pair (i, j) is considered a one-to-one match if $MatchScore(i, j) \geq T_a$, where $T_a = 75\%$.

Using the one-to-one matches (M) identified, along with the number of ground-truth lines (N_1) and detected lines (N_2), the metrics are defined as:

$$\mathrm{DR} = \frac{M}{N_1}, \quad \mathrm{RA} = \frac{M}{N_2}, \quad \mathrm{FM} = \frac{2 \cdot \mathrm{DR} \cdot \mathrm{RA}}{\mathrm{DR} + \mathrm{RA}} \tag{3}$$

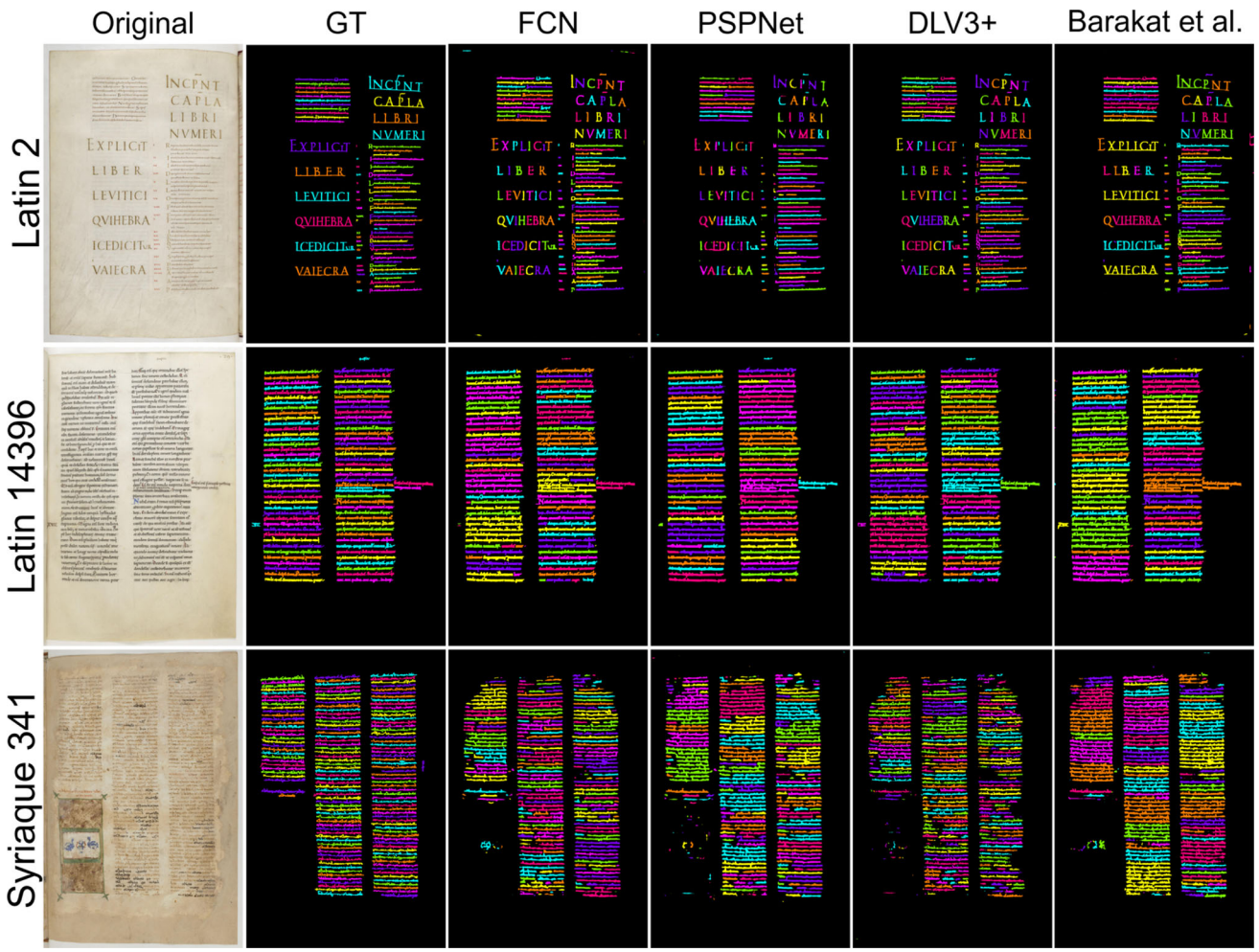


Fig. 5 Qualitative comparison of text line segmentation results. Each color represent a different text line segmentation component

Users are required to follow the original train, validation, and test splits provided with the dataset to ensure comparability with the reported baseline results.

4.2 Results

Table 2 presents the quantitative results of the aforementioned segmentation models on the U-DIADS-TL dataset. The results are provided both for individual manuscripts and for the entire dataset, with scores averaged across each manuscript class.

The DeepLabv3+ model achieved the best performance for Line IU and DR on the full dataset, with an average Line IU of 0.457 and average DR of 0.383. In contrast, FCN and PSPNet showed lower performance, with average Line IU values of 0.366 and 0.314, respectively. Barakat et al., which is a supervised text line approach, achieve the highest overall scores with a RA of 0.312 and FM of 0.322. While DeepLabv3+ performed well, especially with a DR of 0.383, it lagged behind in RA and FM compared to Barakat et al. The results from FCN and PSPNet were consistently lower across all metrics, indicating the challenges these models face in the few-shot text line segmentation setting.

However, even though Barakat et al. achieved better performance compared to the general purpose semantic segmentation models, their performance is still very low on the presented dataset, providing confirmation of the challenge it poses and the need for more advanced techniques to address the challenges of text line segmentation in few-shot settings, especially when dealing with historical and complex manuscripts.

4.3 Sensitivity analysis to different few-shot settings

In Fig. 4, a set of histograms illustrate the performance metrics achieved by the DeepLabv3+ model on each manuscript class of the U-DIADS-TL dataset, when trained in a one-, two-, and three-shot setting. Specifically, for the one-shot setting the performance is reported for each individual image contained in the training set, while for the two and

three-shot settings only the performance achieved for the best possible combination of instances is presented. The query set employed for all few-shot configurations corresponds to the samples identified by the IDs shown in the legend of Fig. 4. As can be observed, the performance improves across all metrics as the number of labeled training images increases. This trend highlights the model's ability to benefit from additional supervision, even in low-resource settings.

However, the relatively poor performance observed in the few-shot setting also underscores the limitations of conventional architectures when faced with extremely scarce training data. These results emphasize the need for dedicated models and techniques specifically designed for one-shot learning in historical document text line segmentation. Given that the creation of ground truth annotations is both time-consuming and costly, especially for ancient and degraded manuscripts, it is crucial to develop approaches that can perform well with minimal supervision.

4.4 Qualitative results

Finally, in Fig. 5 we present a set of qualitative results for the U-DIADS-TL dataset. Specifically, we provide a comparison between the segmentation maps produced by the four aforementioned models and the ground truth maps for each manuscript class that constitutes the dataset. Each color represents a different text line segmentation component. By comparing the segmentations of the approach proposed by Barakat et al. with the GTs, we can observe how the main source of error of the former is represented by its inability to correctly separate the text belonging to subsequent lines effectively, as it tends to group them in a single large connected component. Similarly, it struggles to correctly assign the large letters characterizing the titles of the Latin 2 document class to the same text line. Finally, this approach completely misses some of the text lines present in the GT, while at the same time mistaking some of the graphical elements of the pages as text lines.

5 Conclusion

In this paper, we introduced U-DIADS-TL, a novel dataset specifically designed to advance text line segmentation in historical manuscripts under few-shot learning conditions. Unlike existing datasets, U-DIADS-TL provides noise-free annotations while incorporating diverse and challenging document layouts, including multi-column structures and heterogeneous text orientations. By offering only three training images per document class, this dataset serves as a critical benchmark for developing segmentation models that can generalize effectively with minimal supervision.

Our benchmark results highlight the inherent complexity of historical manuscript analysis, demonstrating that existing segmentation models struggle under few-shot conditions. While the supervised approach by Barakat et al. outperforms generic deep learning architectures, its performance remains limited, reinforcing the need for more robust, data-efficient methods.

Author Contributions S.Z. created the data collection, performed most of the experimentation, and wrote the paper. A.D.N. contributed to writing the paper. C.P. and G.L.F. reviewed the paper.

Funding Open access funding provided by Università degli Studi di Udine within the CRUI-CARE Agreement. Partial financial support was received from Strategic Departmental Plan and interdepartmental center AI4CH – Artificial Intelligence for Cultural Heritage.

Data Availability Data is provided at https://sites.google.com/view/avml-lab-udiadstl/home.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

1. De Nardin, A., Zottin, S., Paier, M., Foresti, G.L., Colombi, E., Piciarelli, C.: Efficient few-shot learning for pixel-precise handwritten document layout analysis. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, Hawaii, 3680–3688 (2023). https://doi.org/10.1109/WACV56688.2023.00367
2. Tarride, S., Lemaitre, A., Coüasnon, B., Tardivel, S.: Combination of deep neural networks and logical rules for record segmentation in historical handwritten registers using few examples. International Journal on Document Analysis and Recognition (IJDAR) **24**(1), 77–96 (2021). https://doi.org/10.1007/s10032-021-00362-8
3. De Nardin, A., Zottin, S., Piciarelli, C., Colombi, E., Foresti, G.L.: Few-shot pixel-precise document layout segmentation via dynamic instance generation and local thresholding. Int. J. Neural Syst. **33**(10), 2350052 (2023). https://doi.org/10.1142/S0129065723500521
4. Zottin, S., De Nardin, A., Foresti, G.L., Colombi, E., Piciarelli, C.: Icdar 2024 competition on few-shot and many-shot layout segmentation of ancient manuscripts (sam). In: Barney Smith, E.H., Liwicki, M., Peng, L. (eds.) Document Analysis and Recognition - ICDAR 2024, pp. 315–331. Springer, Cham (2024)

5. Zottin, S., De Nardin, A., Branca, G., Colombi, E., Piciarelli, C., Shujat, H., Foresti, G.L.: Exploring Few-shot Text Line Segmentation Approaches in Challenging Ancient Manuscripts. In: CEUR Workshop Proceedings, 3937 (2025)
6. De Nardin, A., Zottin, S., Paier, M., Foresti, G.L., Colombi, E., Piciarelli, C.: Dynamic Instance Generation for Few-shot Handwritten Document Layout Segmentation (short Paper). In: CEUR Workshop Proceedings, 3286, 26–34 (2022)
7. Studer, L., Alberti, M., Pondenkandath, V., Goktepe, P., Kolonko, T., Fischer, A., Liwicki, M., Ingold, R.: A comprehensive study of imagenet pre-training for historical document image analysis. In: 2019 International Conference on Document Analysis and Recognition (ICDAR), 720–725 (2019). https://doi.org/10.1109/ICDAR.2019.00120
8. De Nardin, A., Zottin, S., Piciarelli, C., Foresti, G.L., Colombi, E.: In-domain versus out-of-domain transfer learning for document layout analysis. International Journal on Document Analysis and Recognition (IJDAR) (2024). https://doi.org/10.1007/s10032-024-00497-4
9. De Nardin, A., Zottin, S., Colombi, E., Piciarelli, C., Foresti, G.L.: Is imagenet always the best option? an overview on transfer learning strategies for document layout analysis. In: Foresti, G.L., Fusiello, A., Hancock, E. (eds.) Image Analysis and Processing - ICIAP 2023 Workshops, 489–499. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-51026-7_41
10. Barakat, B.K., Droby, A., Alaasam, R., Madi, B., Rabaev, I., Shammes, R., El-Sana, J.: Unsupervised deep learning for text line segmentation. In: 2020 25th International Conference on Pattern Recognition (ICPR), 2304–2311 (2021). https://doi.org/10.1109/ICPR48806.2021.9413308
11. Kurar Barakat, B., Cohen, R., Droby, A., Rabaev, I., El-Sana, J.: Learning-free text line segmentation for historical handwritten documents. Appl. Sci. (2020). https://doi.org/10.3390/app10228276
12. Droby, A., Kurar Barakat, B., Saabni, R., Alaasam, R., Madi, B., El-Sana, J.: Understanding unsupervised deep learning for text line segmentation. Appl. Sci. (2022). https://doi.org/10.3390/app12199528
13. Nikolaidou, K., Seuret, M., Mokayed, H., Liwicki, M.: A survey of historical document image datasets. International Journal on Document Analysis and Recognition (IJDAR) **25**(4), 305–338 (2022). https://doi.org/10.1007/s10032-022-00405-8
14. Simistira, F., Seuret, M., Eichenberger, N., Garz, A., Liwicki, M., Ingold, R.: Diva-hisdb: A precisely annotated large dataset of challenging medieval manuscripts. In: 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), 471–476 (2016). https://doi.org/10.1109/ICFHR.2016.0093
15. Clausner, C., Antonacopoulos, A., Mcgregor, N., Wilson-Nunn, D.: Icfhr 2018 competition on recognition of historical arabic scientific manuscripts – rasm2018. In: 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR), 471–476 (2018). https://doi.org/10.1109/ICFHR-2018.2018.00088
16. Kurar Barakat, B., El-Sana, J., Rabaev, I.: The pinkas dataset. In: 2019 International Conference on Document Analysis and Recognition (ICDAR), 732–737 (2019). https://doi.org/10.1109/ICDAR.2019.00122
17. Potanin, M., Dimitrov, D., Shonenkov, A., Bataev, V., Karachev, D., Novopoltsev, M.: Digital peter: Dataset, competition and handwriting recognition methods. CoRR **abs/2103.09354** (2021) 2103.09354
18. Saini, R., Dobson, D., Morrey, J., Liwicki, M., Simistira Liwicki, F.: Icdar 2019 historical document reading challenge on large structured chinese family records. In: 2019 International Conference on Document Analysis and Recognition (ICDAR), 1499–1504 (2019). https://doi.org/10.1109/ICDAR.2019.00241
19. Alaei, A., Nagabhushan, P., Pal, U.: A new dataset of persian handwritten documents and its segmentation. In: 2011 7th Iranian Conference on Machine Vision and Image Processing, 1–5 (2011). https://doi.org/10.1109/IranianMVIP.2011.6121553
20. Gatos, B., Stamatopoulos, N., Louloudis, G., Sfikas, G., Retsinas, G., Papavassiliou, V., Sunistira, F., Katsouros, V.: Grpoly-db: An old greek polytonic document image database. In: 2015 13th International Conference on Document Analysis and Recognition (ICDAR), 646–650 (2015). https://doi.org/10.1109/ICDAR.2015.7333841
21. Fernández-Mota, D., Almazán, J., Cirera, N., Fornés, A., Lladós, J.: Bh2m: The barcelona historical, handwritten marriages database. In: 2014 22nd International Conference on Pattern Recognition, 256–261 (2014). https://doi.org/10.1109/ICPR.2014.53
22. Kurar Barakat, B., Cohen, R., El-Sana, J.: Vml-moc: Segmenting a multiply oriented and curved handwritten text line dataset. In: 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW), 6, 13–18 (2019). https://doi.org/10.1109/ICDARW.2019.50109
23. Barakat, B.K., Droby, A., Alaasam, R., Madi, B., Rabaev, I., El-Sana, J.: Text line extraction using fully convolutional network and energy minimization. In: Del Bimbo, A., Cucchiara, R., Sclaroff, S., Farinella, G.M., Mei, T., Bertini, M., Escalante, H.J., Vezzani, R. (eds.) Pattern Recognition. ICPR International Workshops and Challenges, 126–140. Springer, Cham (2021)
24. Zottin, S., De Nardin, A., Colombi, E., Piciarelli, C., Pavan, F., Foresti, G.L.: U-diads-bib: a full and few-shot pixel-precise dataset for document layout analysis of ancient manuscripts. Neural Comput. Appl. **36**(20), 11777–11789 (2024). https://doi.org/10.1007/s00521-023-09356-5
25. Zottin, S., De Nardin, A., Branca, G., Piciarelli, C., Foresti, G.L.: Icdar 2025 competition on few-shot text line segmentation of ancient handwritten documents (fest). In: Yin, X.-C., Karatzas, D., Lopresti, D. (eds.) Document Analysis and Recognition - ICDAR 2025, pp. 586–602. Springer, Cham (2026)
26. Sauvola, J., Pietikäinen, M.: Adaptive document image binarization. Pattern Recogn. **33**(2), 225–236 (2000). https://doi.org/10.1016/S0031-3203(99)00055-2
27. De Nardin, A., Zottin, S., Piciarelli, C., Colombi, E., Foresti, G.L.: A one-shot learning approach to document layout segmentation of ancient arabic manuscripts. In: 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 8112–8121 (2024). https://doi.org/10.1109/WACV57701.2024.00794
28. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3431–3440 (2015). https://doi.org/10.1109/CVPR.2015.7298965
29. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 6230–6239 (2017). https://doi.org/10.1109/CVPR.2017.660
30. Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) Computer Vision - ECCV 2018, pp. 833–851. Springer, Cham (2018)
31. Barakat, B., Droby, A., Kassis, M., El-Sana, J.: Text line segmentation for challenging handwritten document images using fully convolutional network. In: 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR), 374–379 (2018). https://doi.org/10.1109/ICFHR-2018.2018.00072
32. Iakubovskii, P.: Segmentation Models Pytorch. GitHub (2019)
33. Gatos, B., Stamatopoulos, N., Louloudis, G.: Icfhr 2010 handwriting segmentation contest. In: 2010 12th International Conference on Frontiers in Handwriting Recognition, 737–742 (2010). https://doi.org/10.1109/ICFHR.2010.120