Exploring few-shot text line segmentation approaches in challenging ancient manuscripts

Silvia Zottin*,[1], Axel De Nardin*,[1], Giuseppe Branca[1], Emanuela Colombi[2], Claudio Piciarelli[1], Hafsa Shujat[3] and Gian Luca Foresti[1]

[1]Department of Mathematics, Computer Science and Physics, University of Udine, Via delle Scienze 206, Udine, Italy
[2]Department of Humanities and Cultural Heritage, University of Udine, Vicolo Florio 2/B, Udine, Italy
[3]International Islamic University Islamabad, Pakistan

Abstract

Text line segmentation is a critical component of document layout analysis, particularly for ancient handwritten manuscripts. Its primary goal is to accurately extract individual text lines, a step that significantly influences subsequent tasks such as optical character recognition, text transcription, and information extraction. However, segmenting text lines in historical manuscripts is particularly challenging due to irregular handwriting, faded ink, and complex layouts with overlapping lines and non-linear text flows. Additionally, the limited availability of large annotated datasets makes fully supervised learning approaches impractical for these documents. In this paper, we explore the applicability of three prominent semantic segmentation models when applied in a few-shot learning setting, using only a small number of labeled examples per manuscript. Our results demonstrate the challenges of addressing text line segmentation in the context of scarce labeled data. This provides a promising avenue for future research in document analysis for historical manuscripts.

Keywords

Text line segmentation, Few-Shot learning, Document layout analysis, Digital manuscript analysis

1. Introduction

Text line segmentation is key in document layout analysis, particularly for historical handwritten manuscripts. This task is critical for downstream applications such as optical character recognition, text transcription, and information extraction. However, segmenting text lines in ancient manuscripts is inherently challenging due to degraded writing, irregular handwriting, complex layouts, overlapping text lines, and non-linear text flows. Additionally, the scarcity of large annotated datasets for these types of documents makes fully supervised learning approaches difficult to implement.

Few-shot learning, which focuses on training models with a limited number of annotated examples, presents a promising solution for addressing these challenges. This approach is particularly valuable for historical manuscripts, where manually labeling large datasets is time-consuming and resource-intensive. In this paper, we explore the application of few-shot learning for text line segmentation in three challenging ancient manuscripts. Specifically, we investigate the performance of three well-known, effective, semantic segmentation models, FCN [1], PSPNet [2], and DeepLabv3+ [3], in the context of few-shot learning. With this paper, we aim to highlight that this is still an unexplored area in the literature and that there is significant potential in developing few-shot techniques for text line segmentation in ancient manuscripts.

The rest of this paper is organized as follows. Section 2 presents the main framework from the existing literature. Section 3 describes the segmentation models employed in our experiments. Section 4

IRCDL 2025: 21st Conference on Information and Research Science Connecting to Digital and Library Science, February 20-21 2025, Udine, Italy

silvia.zottin@uniud.it (S. Zottin*); axel.denardin@uniud.it (A. De Nardin*); branca.giuseppe@spes.uniud.it (G. Branca); emanuela.colombi@uniud.it (E. Colombi); claudio.piciarelli@uniud.it (C. Piciarelli); hafsashujat98@gmail.com (H. Shujat); gianluca.foresti@uniud.it (G. L. Foresti)

0000-0003-0820-7260 (S. Zottin*); 0000-0002-0762-708X (A. De Nardin*); 0000-0002-0384-6664 (E. Colombi); 0000-0001-5305-1520 (C. Piciarelli); 0000-0002-8425-6892 (G. L. Foresti)

*These authors contributed equally to this work provides the details of our experiments and reports the results. Finally, in Section 5, we draw our conclusions and discuss future work.

2. Related Works

The scarcity of extensively labeled data in ancient manuscript analysis is due to the specialized expertise, significant time, and substantial financial resources required to create such datasets, especially for documents with intricate layouts. This limitation naturally motivates the development of systems that can achieve strong performance with minimal annotated data. However, the existing literature offers only a limited number of works that effectively address this challenge.

Unsupervised learning, which does not require any annotated data, has been explored as a potential solution. Most methods in this category rely on intuitive heuristic assumptions that can be translated into deterministic rules. For example, in the task of text line segmentation, a common heuristic assumption is that the document contains only horizontal lines [4, 5]. Assumptions like this one, however, drastically limit the domain of applicability of such systems.

Transfer learning provides another powerful approach to address the data scarcity problem. By leveraging pre-trained deep networks, representations learned on large, general-purpose datasets can be adapted to specific target domains with minimal labeled examples. Research in [6, 7] demonstrates that pre-training on document-related datasets consistently enhances segmentation results and accelerates convergence compared to using general-purpose datasets. These findings underscore the benefits of domain-specific pre-training for specialized applications. In addition, the study in [8] investigates the performance of models pre-trained on ImageNet and fine-tuned on an ancient document dataset. It highlights that the effectiveness of transfer learning versus training from scratch depends heavily on the characteristics of the target dataset. Similarly, domain-specific transfer learning has been shown to improve performance in document layout segmentation tasks.

Few-shot and one-shot learning techniques have also emerged as promising solutions for data-scarce scenarios. An example of a few-shot learning strategy is presented in [9], where the model is trained with only two labeled images per manuscript. This framework combines a novel data augmentation technique [10] with a segmentation refinement module based on a traditional local thresholding method [11], achieving results comparable to state-of-the-art supervised methods. Another few-shot technique, Deep & Syntax [12], focuses on segmenting historical handwritten registers by leveraging recurrent patterns to delineate individual records. This hybrid approach combines U-shaped neural networks with logical rules, such as filtering and text alignment, to enhance accuracy.

A more recent one-shot learning framework is introduced in [13], targeting layout segmentation of ancient Arabic documents. Despite using only one labeled page per manuscript, this method achieves state-of-the-art performance on a challenging dataset. It incorporates three main components: a semantic segmentation backbone, a dynamic instance generation module, and a segmentation refinement module. These advancements emphasize the growing interest in addressing the challenges of low-data scenarios. The significance of few-shot learning is further highlighted by the SAM challenge [14], which focuses on document layout segmentation under few-shot conditions.

Despite these developments, few-shot learning techniques have not yet been explored for text line segmentation. To the best of our knowledge, no studies have addressed this specific topic in the literature. In this paper, we aim to fill this gap by exploring few-shot learning for text line segmentation in ancient manuscripts. We present preliminary results demonstrating the feasibility of performing text line segmentation using only three labeled images per manuscript.

3. Methods

In this paper, we explore the performance of three prominent and high-performing semantic segmentation models when adopted for text line segmentation of ancient manuscripts in a few-shot setting.

Specifically, we focus on three of the most popular models in the literature characterized by different architectural choices: FCN [1], PSPNet [2], and DeepLabv3+ [3].

3.1. Fully Convolutional Network

Fully Convolutional Network (FCN) [1] is a pioneering architecture designed for semantic segmentation tasks. The core architecture of FCN builds upon standard convolutional neural networks by transforming them into end-to-end trainable models capable of dense predictions. In particular, fully connected layers are replaced with fully convolutional layers, preserving spatial information and allowing the network to process input images of arbitrary size. To recover spatial resolution lost during downsampling in convolutional and pooling layers, FCN introduces upsampling layers. These layers use deconvolution to produce predictions at the same resolution as the input image.

FCN is characterized by an encoder-decoder structure. The encoder comprises convolutional and pooling layers from standard classification networks, such as VGG or ResNet. These layers extract hierarchical features by progressively reducing spatial resolution while capturing high-level semantic information. The decoder restores the spatial resolution of the feature maps to match the input image dimensions. This is achieved through upsampling layers, which use learned deconvolution filters to interpolate the feature maps back to the original resolution. Skip connections are introduced to combine lower-level features from the encoder with higher-level features in the decoder, improving localization and boundary accuracy.

One of the key innovations in FCN is the use of skip connections, which fuse feature maps from different layers. These connections combine coarse semantic information from deeper layers with fine spatial details from shallower layers, leading to more accurate segmentation, especially at object boundaries.

3.2. Pyramid Scene Parsing Network

Pyramid Scene Parsing Network (PSPNet) [2] is a powerful model for semantic segmentation, known for its ability to capture both local and global contextual information. By employing a pyramid pooling module, PSPNet excels in understanding complex scenes with objects of varying sizes and arrangements. PSPNet builds upon a backbone network, typically ResNet [15], which serves as the feature extractor. The core innovation of PSPNet lies in the Pyramid Pooling Module (PPM), which aggregates contextual information from multiple spatial scales.

The encoder extracts hierarchical feature maps from the input image. A pre-trained ResNet, truncated before the fully connected layers, is commonly used. This backbone captures rich semantic features at reduced spatial resolution through convolutional and pooling operations.

The PPM is the centerpiece of PSPNet, designed to enhance the feature representation by pooling information from different regions of the image. Feature maps from the encoder are pooled into different grid sizes (e.g., 1×1, 2×2, 3×3, 6×6). Each pooled representation captures contextual information from increasingly larger regions, ranging from global to local. Finally, the pooled outputs are upsampled to match the resolution of the original feature map and concatenated with it, resulting in a feature map enriched with multi-scale context.

The decoder takes the enriched feature map from the PPM and refines it through convolutional layers. This process generates pixel-wise predictions for the semantic segmentation task, with an output resolution matching the input image.

3.3. DeepLabv3+

DeepLabv3+ [3] is a state-of-the-art model for semantic segmentation, designed to achieve high accuracy by effectively capturing contextual information and refining spatial details. It builds upon the DeepLabv3 [16] architecture by incorporating an encoder-decoder structure, enabling more precise boundary delineation in segmentation tasks. Below, we provide a detailed description of the key components and mechanisms of DeepLabv3+.

The encoder in DeepLabv3+ employs atrous (dilated) convolutions to expand the receptive field of convolutional layers without increasing the number of parameters or losing spatial resolution. Atrous convolutions allow the model to capture multi-scale context efficiently, which is crucial for recognizing objects at various scales.

To further enhance multi-scale feature extraction, the encoder relies on the Atrous Spatial Pyramid Pooling (ASPP) module. ASPP applies parallel atrous convolutions with different dilation rates, capturing context at multiple resolutions. In addition to atrous convolutions, ASPP includes image-level pooling to aggregate global contextual information. The resulting feature maps are concatenated and processed through a 1x1 convolution layer and a batch normalization layer, producing robust multi-scale features.

While the encoder captures high-level semantic information, it often results in reduced spatial resolution due to downsampling operations. To address this, DeepLabv3+ introduces a decoder module that refines the segmentation output by recovering spatial details.

The decoder upsamples the output of the encoder using bilinear interpolation, aligning it with the spatial resolution of the lower-level feature maps extracted from earlier stages of the network. These lower-level feature maps, which retain finer spatial information, are then concatenated with the upsampled encoder output. This fused representation is further processed through convolutional layers, producing sharper and more accurate segmentation results, particularly around object boundaries.

4. Experiments

4.1. Dataset

The dataset used in this study is a proprietary collection developed through a collaborative effort between computer scientists and humanities scholars. It was derived by processing images from the publicly available U-DIADS-Bib dataset [17], primarily focusing on document layout segmentation.

Our dataset consists of 105 images, divided into 35 images for each of three distinct ancient manuscripts: Latin 2, Latin 14396, and Syriaque 341. Each manuscript varies in layout structure, degradation levels due to preservation and aging, and the alphabet in which it is written. Specifically, Latin 2 is a Latin-language manuscript arranged in two columns and featuring various interlinear paratexts. Similarly, Latin 14396, also in Latin, consists of text in two columns but includes diverse fonts and marginal paratexts, making its segmentation more variable. Finally, Syriaque 341 is a Syriac-language manuscript arranged in three columns. This manuscript is the most challenging, as it includes vertical comments and paratexts, and its pages are highly degraded due to aging and poor preservation. These unique characteristics of each manuscripts make the dataset particularly intriguing and challenging. The images of the three manuscripts were sourced from the French digital library Gallica[1].

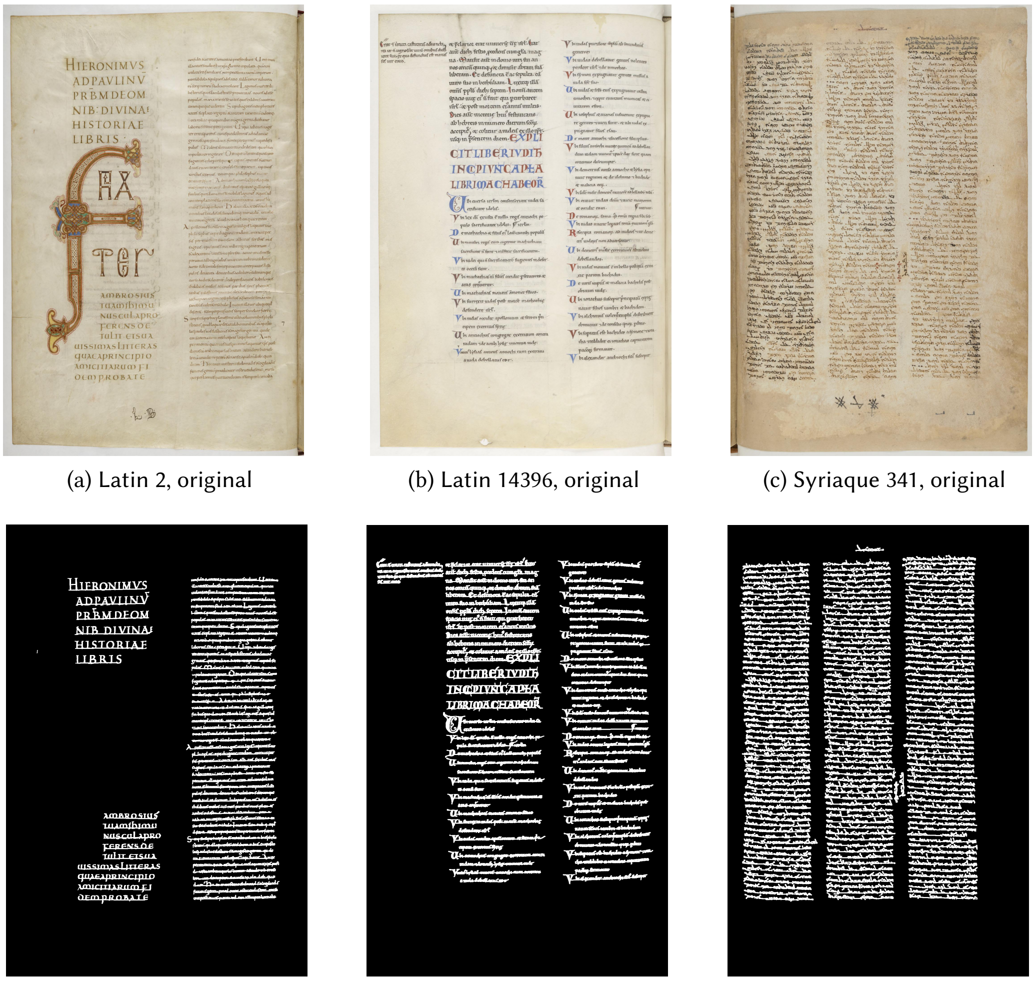
Given that the dataset was designed for use with few-shot learning techniques, 35 unique color page images were created for each manuscript, divided into three sets: 3 images for training, 10 for validation, and 15 for testing. Each page is paired with corresponding Ground Truth (GT) data, which includes two distinct and non-overlapping annotated classes: background and text lines.

Figure 1 illustrates examples of the defined GT and their corresponding original images for each manuscript. Unlike many currently available document text line segmentation datasets, our proprietary dataset is characterized by pixel-level precision, non-overlapping elements, the absence of noise, heterogeneously oriented text lines, and a multi-column layout.

4.2. Evaluation Metrics

Evaluation involves calculating key text line semantic segmentation metrics commonly adopted in the literature [18, 19], including Pixel Intersection over Union (Pixel IU), Line Intersection over Union (Line IU), Detection Rate (DR), Recognition Accuracy (RA), and F-measure (FM).

[1] Source: https://gallica.bnf.fr



(a) Latin 2, original

(b) Latin 14396, original

(c) Syriaque 341, original

(d) Latin 2, GT

(e) Latin 14396, GT

(f) Syriaque 341, GT

Figure 1: An example page for each manuscript that composes the dataset used for training and testing (a - c) and the corresponding GT masks (d - f). The background is highlighted in black, while each text line that composes the manuscript page is highlighted in white with pixel-level precision.

Pixel IU and Line IU are 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 denotes True Positives, FP False Positives, and FN False Negatives. Pixel IU evaluates IU at the pixel level, measuring the accuracy of line detection. Here, TP represents correctly detected pixels, FP represents extra (false) pixels, and FN represents missed pixels. Line IU evaluates IU at the line level, measuring how accurately lines are detected. In this case, TP is the number of correctly detected lines, FP is the number of extra lines, and FN is the number of missed lines. A threshold of 75% is applied to determine matches between predicted and ground-truth connected components. Two components 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).

Detection Rate (DR), Recognition Accuracy (RA), and F-Measure (FM) rely on the MatchScore metric. For a given image, let R_i represent 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 calculated 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 $\text{MatchScore}(i, j) \geq T_a$, 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:

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

These metrics are calculated for each manuscript individually, with the overall dataset average provided for each metric as well.

4.3. Hyper-parameters setup

For training the proposed model, the Jaccard loss function was selected. Jaccard loss quantifies the dissimilarity between the predicted segmentation mask and the GT mask, leveraging the intersection and union of the two masks. The Jaccard loss is defined as:

$$\mathcal{L}_{\text{Jaccard}} = 1 - \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}, \tag{4}$$

where TP, FP, and FN represent true positives, false positives, and false negatives, respectively.

The model was optimized using the ADAM optimizer with a learning rate of 1×10^{-3} and a weight decay of 1×10^{-5}. Training was conducted for a maximum of 100 epochs, with an early stopping criterion applied if the network showed no improvement over the last 20 iterations after completing 50 epochs. For all models, a ResNet50 was chosen as the backbone.

4.4. Results

Table 1 summarizes the performance evaluation of the three semantic segmentation models, FCN [1], PSPNet [2], and DeepLabv3+ [3], on the text line segmentation task. The results are reported both for individual manuscripts and for the entire dataset, with averages calculated across the three manuscript classes: Latin 2, Latin 14396, and Syriaque 341.

Among the evaluated models, DeepLabv3+ shows the best overall performance across all metrics when considering the average scores for the entire dataset. Specifically, it achieves the highest Pixel IU (0.490), Line IU (0.443), DR (0.291), RA (0.177), and FM (0.217). FCN exhibits competitive performance on the Latin 2 and Latin 14396 manuscripts, with an average Pixel IU of 0.416 and Line IU of 0.340. However, it struggles with the more challenging Syriaque 341 manuscript, where degradation and complex layouts significantly affect its performance. PSPNet, while demonstrating reasonable results on Latin 14396, performs less consistently overall, particularly on the Syriaque 341 manuscript, with an average Pixel IU of 0.365 and Line IU of 0.314.

Overall, the results are notably low, which highlights the significant challenges still present in the task of text line segmentation for historical documents in a few-shot setting.

5. Conclusion

In this study, we explored few-shot learning approaches for text line segmentation in ancient manuscripts, a challenging task due to the unique characteristics of historical documents. We applied three semantic segmentation models and evaluated their performance on a proprietary dataset of ancient manuscripts, consisting of Latin and Syriac texts with varying levels of degradation and complexity.

The results revealed that while DeepLabv3+ achieved the best overall performance, the performance across all models was relatively low, indicating the inherent difficulties of segmenting text lines in such challenging datasets. Despite this, the experiments demonstrate that few-shot learning can be a

Backbone	Metrics	Latin 2	Latin 14396	Syriaque 341	Average
FCN [1]	Pixel IU	0.513	0.551	0.183	0.416
	Line IU	0.449	0.493	0.079	0.340
	DR	0.292	0.437	0.030	0.253
	RA	0.166	0.269	0.041	0.159
	FM	0.210	0.329	0.034	0.191
PSPNet [2]	Pixel IU	0.515	0.503	0.079	0.365
	Line IU	0.402	0.519	0.020	0.314
	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+ [3]	Pixel IU	0.554	0.573	0.342	**0.490**
	Line IU	0.530	0.582	0.218	**0.443**
	DR	0.275	0.512	0.086	**0.291**
	RA	0.148	0.331	0.051	**0.177**
	FM	0.191	0.397	0.064	**0.217**

Table 1
Performance comparison of the different semantic segmentation models chosen tested on our dataset. The best-performing results for full dataset are shown in bold.

promising approach for text line segmentation, offering a viable solution in situations where annotated data is scarce.

With this research, we highlight the need and the possibility of more effective methods for the segmentation of ancient manuscripts, contributing to the preservation and accessibility of cultural heritage.

Acknowledgments

Partial financial support was received from: PNRR DD 3277 del 30 dicembre 2021 (PNRR Missione 4, Componente 2, Investimento 1.5) - iNEST; from Strategic Departmental Plan on Artificial Intelligence, DMIF, University of Udine; from Strategic Departmental Plan and interdepartmental center AI4CH – Artificial Intelligence for Cultural Heritage of University of Udine.

References

[1] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3431–3440. doi:10.1109/CVPR.2015.7298965.

[2] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6230–6239. doi:10.1109/CVPR.2017.660.

[3] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: V. Ferrari, M. Hebert, C. Sminchisescu, Y. Weiss (Eds.), Computer Vision – ECCV 2018, Springer International Publishing, Cham, 2018, pp. 833–851.

[4] Z. Li, W. Wang, Y. Chen, Y. Hao, A novel method of text line segmentation for historical document

image of the uchen tibetan, Journal of Visual Communication and Image Representation 61 (2019) 23–32. URL: https://www.sciencedirect.com/science/article/pii/S1047320319300288. doi:https://doi.org/10.1016/j.jvcir.2019.01.021.

[5] T.-N. Nguyen, J.-C. Burie, T.-L. Le, A.-V. Schweyer, An effective method for text line segmentation in historical document images, in: 2022 26th International Conference on Pattern Recognition (ICPR), IEEE, Montreal, Canada, 2022, pp. 1593–1599. URL: https://hal.science/hal-03922470. doi:10.1109/ICPR56361.2022.9956617.

[6] A. De Nardin, S. Zottin, E. Colombi, C. Piciarelli, G. L. Foresti, Is imagenet always the best option? an overview on transfer learning strategies for document layout analysis, in: G. L. Foresti, A. Fusiello, E. Hancock (Eds.), Image Analysis and Processing - ICIAP 2023 Workshops, Springer Nature Switzerland, Cham, 2024, pp. 489–499. doi:https://doi.org/10.1007/978-3-031-51026-7_41.

[7] A. De Nardin, S. Zottin, C. Piciarelli, G. L. Foresti, E. Colombi, In-domain versus out-of-domain transfer learning for document layout analysis, International Journal on Document Analysis and Recognition (IJDAR) (2024). URL: https://doi.org/10.1007/s10032-024-00497-4. doi:10.1007/s10032-024-00497-4.

[8] L. Studer, M. Alberti, V. Pondenkandath, P. Goktepe, T. Kolonko, A. Fischer, M. Liwicki, R. Ingold, A comprehensive study of imagenet pre-training for historical document image analysis, in: 2019 International Conference on Document Analysis and Recognition (ICDAR), 2019, pp. 720–725. doi:10.1109/ICDAR.2019.00120.

[9] A. De Nardin, S. Zottin, M. Paier, G. L. Foresti, E. Colombi, C. Piciarelli, 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, 2023, pp. 3680–3688. doi:10.1109/WACV56688.2023.00367.

[10] A. De Nardin, S. Zottin, M. Paier, G. L. Foresti, E. Colombi, C. Piciarelli, Dynamic instance generation for few-shot handwritten document layout segmentation (short paper), in: CEUR Workshop Proceedings, volume 3286, 2022, p. 26 – 34.

[11] A. De Nardin, S. Zottin, C. Piciarelli, E. Colombi, G. L. Foresti, Few-shot pixel-precise document layout segmentation via dynamic instance generation and local thresholding, International Journal of Neural Systems 33 (2023) 2350052. doi:10.1142/S0129065723500521.

[12] S. Tarride, A. Lemaitre, B. Coüasnon, S. Tardivel, 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 (2021) 77–96. doi:10.1007/s10032-021-00362-8.

[13] A. De Nardin, S. Zottin, C. Piciarelli, E. Colombi, G. L. Foresti, 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), 2024, pp. 8112–8121. doi:10.1109/WACV57701.2024.00794.

[14] S. Zottin, A. De Nardin, G. L. Foresti, E. Colombi, C. Piciarelli, Icdar 2024 competition on few-shot and many-shot layout segmentation of ancient manuscripts (sam), in: E. H. Barney Smith, M. Liwicki, L. Peng (Eds.), Document Analysis and Recognition - ICDAR 2024, Springer Nature Switzerland, Cham, 2024, pp. 315–331.

[15] Z. Cheng, F. Bai, Y. Xu, G. Zheng, S. Pu, S. Zhou, Focusing attention: Towards accurate text recognition in natural images, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017.

[16] L. Chen, G. Papandreou, F. Schroff, H. Adam, Rethinking atrous convolution for semantic image segmentation, CoRR abs/1706.05587 (2017). arXiv:1706.05587.

[17] S. Zottin, A. De Nardin, E. Colombi, C. Piciarelli, F. Pavan, G. L. Foresti, U-diads-bib: a full and few-shot pixel-precise dataset for document layout analysis of ancient manuscripts, Neural Computing and Applications 36 (2024) 11777–11789. URL: https://doi.org/10.1007/s00521-023-09356-5. doi:10.1007/s00521-023-09356-5.

[18] F. Simistira, M. Seuret, N. Eichenberger, A. Garz, M. Liwicki, R. Ingold, Diva-hisdb: A pre-

cisely annotated large dataset of challenging medieval manuscripts, in: 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2016, pp. 471–476. doi:10.1109/ICFHR.2016.0093.

[19] M. W. A. Kesiman, D. Valy, J. C. Burie, E. Paulus, M. Suryani, S. Hadi, M. Verleysen, S. Chhun, J.-M. Ogier, Icfhr 2018 competition on document image analysis tasks for southeast asian palm leaf manuscripts, in: 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2018, pp. 483–488. doi:10.1109/ICFHR-2018.2018.00090.