SeamFormer: High Precision Text Line Segmentation for Handwritten Documents

Niharika Vadlamudi[0000−0001−6596−428X], Rahul Krishna[0000−0002−8734−5751],
and Ravi Kiran Sarvadevabhatla[0000−0003−4134−1154]

Centre for Visual Information Technology International Institute of Information Technology, Hyderabad – 500032, INDIA
ravi.kiran@iiit.ac.in

Abstract Historical palm leaf manuscripts often contain dense unstructured text lines. The large diversity in sizes, scripts and appearance makes precise text line segmentation extremely challenging. Existing line segmentation approaches often associate diacritic elements incorrectly to text lines and also address above mentioned challenges inadequately. To tackle these issues, we introduce SeamFormer, a novel approach for high precision text line segmentation in handwritten manuscripts. In the first stage of our approach, a multi-task Transformer deep network outputs coarse line identifiers which we term 'scribbles' and the binarized manuscript image. In the second stage, a scribble-conditioned seam generation procedure utilizes outputs from first stage and feature maps derived from manuscript image to generate tight-fitting line segmentation polygons. In the process, we incorporate a novel diacritic feature map which enables improved diacritic and text line associations. Via experiments and evaluations on new and existing challenging palm leaf manuscript datasets, we show that SeamFormer outperforms competing approaches and generates precise text line segmentations.

Keywords: Text Line Segmentation · Historical Manuscripts

1 Introduction

Identifying text lines in ancient handwritten documents is an important problem in document image understanding [53,14,8,15,17,26]. Since historical documents usually contain text written in a highly unstructured manner with dense and non-standard layouts, the problem is challenging. The challenge aspect is particularly apparent for palm leaf manuscripts of South-East Asia and the Indian subcontinent. Western manuscripts predominantly use processed animal-skin (vellum) as their base material. Though these are not immune to ravages of time, palm leaf manuscripts are relatively more fragile. Also, palm leaves are thin, delicate and prone to damage. Moreover, the already faintly written text may fade over time and become indistinguishable from digitization noise. Document analysis tasks on palm leaf manuscripts involve characteristic challenges such as degradation, low contrast, variable inter-character and inter-line spacing

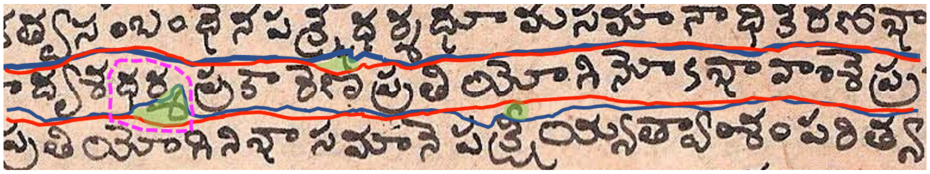

Figure 1. An example to illustrate the importance of precise line segmentation in palm leaf manuscripts. The ground truth upper and lower portions of the enclosing line annotation are shown in red. The prediction is shown in blue. The green shaded portions indicate crucial text fragments omitted by prediction causing the semantic interpretation of text to change. For e.g., pink dashed region encloses a word ధర్మ from the Indic language Telugu which means 'moral duty'. The incorrect boundary prediction causes the resulting line to contain a word ధర with a drastically different meaning. ధర means 'price'.

and morphological distortions in character shapes [38,44,25]. The large diversity in spatial dimensions, languages, scripts, writing styles and presence of non-textual elements further compound the challenge for text line segmentation.

The output of text line segmentation is often processed by a subsequent Optical Character Recognition (OCR) module. Obtaining high precision segmentation maps of text lines which could be used as masks compactly enclosing the reference text is extremely crucial. Using such masks within the OCR pipeline reduces semantic noise from adjoining line fragments generally present in the text-line's bounding box and potentially increases OCR performance. Indic and South-East Asian manuscript texts are characterized by orthographic text fragments such as diacritics. These components typically exist at varying distances from the parent text line. Due to the semantics associated with such components, omission or incorrect association of diacritics to text lines during segmentation can result in a dramatically modified linguistic interpretation of the text (see Fig. 1). Therefore, it is essential to develop segmentation approaches for palm leaf manuscripts which are highly precise. The performance of existing line segmentation approaches fall short in this aspect.

To tackle the challenge, we propose SeamFormer, a robust text line segmentation framework for palm leaf manuscripts. SeamFormer is configured as a two stage pipeline (Sec. 3). In the first stage, the manuscript image is processed by a multi-task Transformer deep network to obtain the binarized image and coarse identifiers for each text line which we term 'scribbles' (Sec. 3.1). In the second stage (Sec. 3.2), the extracted scribbles, binarized image and custom-designed feature maps are fed to a scribble-conditioned seam generation algorithm which generates the desired tight fitting polygons enclosing the individual text lines. Via experiments and evaluations on new and existing palm leaf manuscript datasets, we show that SeamFormer generates significantly superior line segmentations compared to other competing approaches (Sec. 5).

The source code, pretrained models and associated material are available at this link: https://ihdia.iiit.ac.in/seamformer.

2 Related Work

Many approaches have been proposed for text line segmentation in other (i.e. non palm leaf) historical documents. To encourage research, many historical document datasets with line segmentation annotations have been introduced and utilized in competitions at premier document analysis venues - refer to the comprehensive survey paper by Nikolaidou et al. [34] for details.

Early approaches favoured the use of classical digital image processing techniques followed by post processing. Alaei et al. [1] employ a painting technique for foregrounding smearing to tackle unconstrained handwritten text line segmentation for diverse languages. Grouping techniques utilizing nearest neighbor [35], learning algorithms [39], and heuristic rules [28] have also been employed for text line segmentation. Projection profiles are another popular top-down approach to isolate text lines [10,19,31,37,54]. However, profile-based approaches cannot cope with highly curved lines and uneven layouts. Adaptive Local Connectivity Map (ALCM) [45,46] is another technique for localizing and extracting text lines directly from gray-scale images. Generally, these approaches employ handcrafted processing elements with hyperparameters which do not generalize well across multiple datasets. The methods tend to require dataset specific techniques for isolating text line elements (e.g. strokes, diacritics) and often fail to disentangle touching components across consecutive text lines – a common occurrence in handwritten documents.

In recent years, a number of deep learning based approaches have been employed as well [40,7,6,9,29,36,30,27]. Most of these approaches use a variant of the popular U-Net [41] architecture. These methods have the appeal of being optimized end-to-end and work well on Western historical manuscripts. However, the approaches require drastic downsampling of input image which eliminates crucial inter-line information. Coupled with the boundary smoothing that occurs when the network predictions are upsampled, this leads to imprecise and unsatisfactory line segment boundary predictions for other types of historical manuscripts such as ours (i.e. palm leaf).

Relatively few works have tackled line segmentation for palm leaf manuscripts. In their survey paper, Kesiman et al. [25] consider palm-leaf manuscripts from South-East Asia and evaluate numerous line segmentation approaches developed for other (non-Asian) historical documents. Chamchong and Fung propose an adaptive partial projection (APP) technique [13], an improvement over their earlier partial projection approach [12] for line extraction in Thai manuscripts. Valy et al. [51] propose an approach which also employs connected components and projection profiles to determine medial positions of text lines followed by a path finding approach to mark the text line boundaries in Khmer manuscripts. Kesiman et al. [22] employ a similar approach for Balinese manuscripts. Apart from the assumption of a component-based script, these approaches inherit the shortcomings of projection-based works mentioned previously.

Works which employ deep neural networks for palm leaf manuscript text line segmentation are even fewer. Jindal and Ghosh [21] use a Faster-RCNN model to obtain bounding boxes for a collection of Indic palm leaf manuscripts. However, this approach cannot tackle the curvature of lines which is present in almost all manuscripts. Prusty et al. [38] and Sharan et al. [44] propose approaches which modify the Mask-RCNN [20] framework for segmenting various semantic regions including text lines in Indic manuscripts. Despite their relatively better performance and ability to tackle line curvature, these approaches produce overly smoothed line boundaries and even tend to have false negatives (i.e. missed lines) on some occasions.

Seam generation, an approach involving optimization over image-derived energy maps [5], is a popular approach for text line segmentation. Saabni and El-Sana introduce a seam generation algorithm based on an energy map calculated using Signed Distance Transform (SDT) for Arabic manuscripts [42]. However, the approach involves repeated energy map computations for each line and significant dataset-specific post-processing to tackle overlapping components and diacritics. Asi et al. [4] improve upon the aforementioned approach by replacing SDT with a geodesic distance transform energy map. This method fails to tackle elongated letters and widely separated diacritics. Nikolaos et al. [3] use a medial line obtained using a projection profile approach to guide seam generation for line segmentation in multiple historical datasets. However, the method requires dataset specific parameter tuning for various pipeline stages. Alberti et al. [2] first employ a deep network to obtain a binarized version of the image. Seam generation is applied on the binary image to obtain coarse region boundaries, followed by a graph-based connected component procedure to obtain the polygonal line boundaries. The approach is not suitable for highly skewed and unevenly curving text found in palm leaf manuscripts.Nguyen et al. [32] also apply seam carving approach to binarised images. For enhancing the seam generation process , along with the energy map they have introduced a global cost function for better detection of the ascenders, descenders and diacritics. The approach is not suitable for skewed or curved text and requires heavy dataset specific parameter tuning for its proposed cost functions.

In existing approaches [42,4,3,2,27], seam generation is generally used to separate text lines rather than *segment* them. As a result, extraneous isolated character fragments and noisy background elements present beyond the line's text content are often included as part of the line. In contrast, our approach generates polygons which compactly enclose the text lines. As a novel element, we introduce a custom energy map in our polygon generation stage which emphasizes proper association of diacritics to the parent text line. Another marked departure from existing methods is the absence of final post-processing. This enables our approach to generalize across multiple palm leaf manuscript datasets containing documents with varying scripts and text line densities.

3 Approach

Overview: Given the input palm leaf manuscript image, our objective is to generate tight-fitting polygons enclosing each of the text lines. Our processing pipeline has two stages – 'scribble generation' (Sec. 3.1) and 'text line polygon

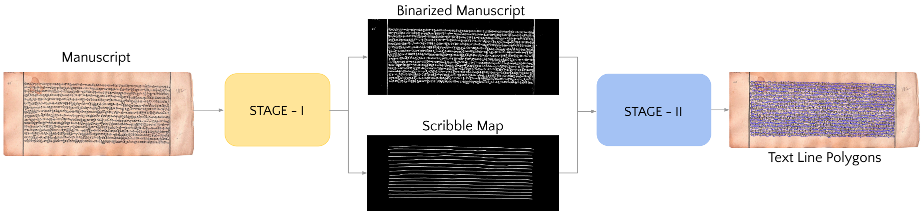

Figure 2. An outline of our SeamFormer pipeline for manuscript line segmentation (Sec. 3.

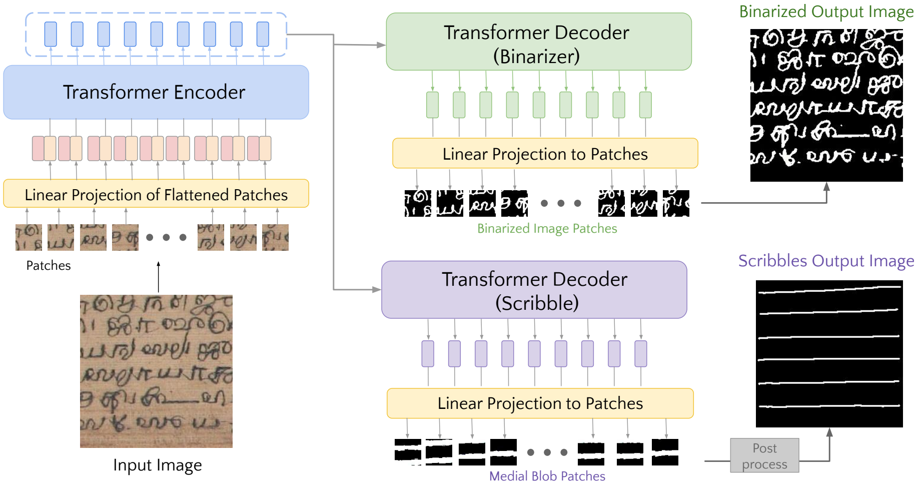
Transformer Encoder
Linear Projection of Flattened Patches
Patches
Input Image
Transformer Decoder (Binarizer)
Linear Projection to Patches
Binarized Image Patches
Binarized Output Image
Transformer Decoder (Scribble)
Linear Projection to Patches
Medial Blob Patches
Post process
Scribbles Output Image

Figure 3. Stage I: Scribble Generation Module - see Sec. 3.1.

generation' (Sec. 3.2) – see Fig. 2. In the first stage, the manuscript image is processed by a deep network which generates coarse binary medial blobs for each individual text line and a binarized version of the image. The medial blobs are further processed to extract coarse spatial identifiers for each line termed as 'scribbles'. In the second stage, scribbles from first stage and custom-designed feature maps derived from binarized image are fed to a seam generation algorithm which generates the desired tight-fitting polygons enclosing the individual text lines.

3.1 Stage I: Scribble Generation

We set up a multi-task variant of Vision Transformer (ViT) deep network architecture [16] to obtain two outputs - the binarized version of the input manuscript image and the medial blob masks for each text line (see Fig. 3). In a conventional ViT architecture, position-encoded patches of input image are processed within a Transformer [52] framework employing multi-head attention to obtain output

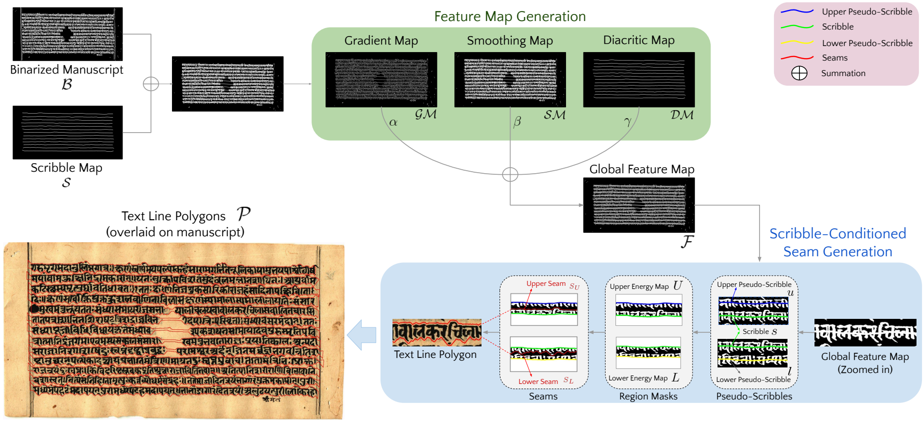

Figure 4. Stage II: Text Line Polygon Generation Module - see Sec. 3.2 and Algorithm 1 for details.

patches. We extend the conventional setup to have two decoder branches. These branches output two sets of patches which are separately reassembled to obtain the binarized version of the input image and the medial blob masks binary image.

The blob mask outputs are post-processed to extract thin medial axis-like structures which cut across the line. We broadly classify our post-processing into local and global stages. In local post processing, we iteratively apply morphological dilation and erosion on each blob mask and perform skeletonisation. Subsequently, we apply skeleton pruning techniques to remove spurious branches and extract a clean medial fragment for each blob within the patch. We term these fragments as 'scribbles'. For the global post processing, we merge these patches to obtain a scribble map with the input image's dimensions. Given the fragments of scribbles, we group them based on distance thresholding technique as a function of its horizontal level.

The scribble, by nature of its construction, provides crucial information regarding local curvature of the text line. As we shall see, accurate determination of local curvature plays a key role for the next stage of processing and ultimately, for accurate text line segmentation.

3.2 Stage II: Text Line Polygon Generation

This stage involves two sub-stages – Feature Map Generation and Scribble-conditioned Seam Generation (see Fig. 4). For each scribble, we first generate a corresponding pair of *pseudo-scribbles* which are used at later stages of the pipeline (Sec. 3.2.1). Next, the scribbles are overlaid on binarized input image and the resulting scribble-overlaid image is used to create custom feature maps (Sec. 3.2.2). These feature maps are used as input to a seam generation procedure to generate the desired high-precision polygons enclosing the text lines (Sec. 3.2.3).

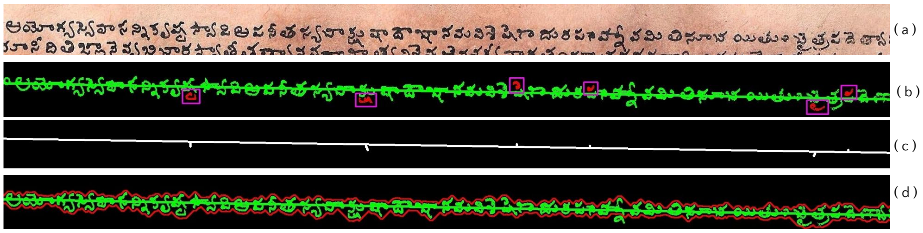

Figure 5. Diacritic Map (Sec. 3.2.2) - (a) A text line from a palm leaf manuscript, (b) the reference text line is shown with the scribble overlaid. Pixels in green denote the text line connected by the scribble and pixels in red inside pink bounding boxes denote the corresponding diacritics of the parent text line (c) Diacritic Feature Map - note the tiny strokes extending out of the scribble to connect the diacritics with the main text line (d) final red seams enclosing the text line as a result of using the Diacritic map during seam generation - note that the aforementioned diacritics have been brought inside the enclosing seams.

3.2.1 Pseudo-scribble generation

As the first step, we sort the scribbles by the y-coordinate of the left-most point to obtain the sequence of scribbles S in a top-to-bottom order. Let $s_i \in \mathcal{S}$ be a scribble. Let μ_{s_i} be the average of all y-coordinates of the scribble s_i's pixels. Let $\mu_{s_{i+1}}$ be a similar average for the neighboring scribble. The vertical offset between the scribble pair can be defined as $d(s_i, s_{i+1}) = \left|\mu_{s_i} - \mu_{s_{i+1}}\right|$. Let $\overline{d(\mathcal{S})}$ denote the average across all such vertical offsets within the set of scribbles. Define $\theta = \overline{d(\mathcal{S})} + \delta$ where δ is a fixed offset. For each scribble s, the upper pseudo-scribble (u) and lower pseudo-scribble (l), are obtained by vertically translating s by $+\theta$ and $-\theta$ pixels respectively – see the block 'Pseudo scribbles' which is part of 'Scribble-Conditioned Seam-Generation' (shaded blue) in Fig. 4.

3.2.2 Feature Map Generation

Gradient Map (GM): This feature map is obtained as the gradient magnitude map of the scribble-overlaid image. Using this map creates a high energy barrier between edges of characters in the text line and the background area immediately surrounding them. Employing this map in the subsequent seam generation stage enables seams to align closely with text letter boundaries, resulting in tight-fitting polygons around the text lines (ref. $\mathcal{GM}$ in Fig. 4).
Smoothing Map (SM): This feature map is obtained by applying a blur kernel on the scribble-overlaid image. Using this map increases the energy at horizontal inter-character text gaps and ensures that seams do not cut through the text (ref. $\mathcal{SM}$ in Fig. 4).
Diacritic Map (DM): This novel feature map specifically tackles the problem of diacritics not being enclosed within the polygons of corresponding parent text

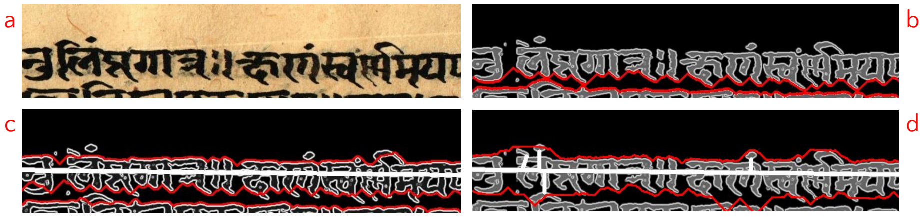

Figure 6. (a) A fragment from the top portion of a manuscript (b) Seams generated with Gradient and Smoothing Map, but without using scribble – the upper line boundary is missing (c) Seams when scribble is also added – upper line boundary is obtained, but diacritics are missed (d) Seams when Diacritic Map is also included – line boundaries properly enclose text and associated diacritic components.

lines - see Fig. 5. We first isolate the region around each text line with the help of upper and lower pseudo-scribbles as the demarcations. We overlay the corresponding scribble on the parent text-line and perform connected components analysis. This operation divides the components into three major groups: components connected to parent-line, disconnected diacritics and background noisy elements. We discard noise based on an area threshold. For each diacritic component, we connect its centroid and parent scribble via a perpendicular line. In effect, this line creates an energy barrier which forces the boundary generated during seam generation to move around the diacritic instead of separating the diacritic and its parent text line (ref. $\mathcal{DM}$ in Fig. 4). The utility of Diacritic Map is illustrated in Fig. 5. The neighborhood of a text line often contains text fragments from adjacent lines due to the uneven handwritten line orientation and dense handwriting. Our construction of the Diacritic Map actively prevents the neighbouring text fragments from being picked up along with the diacritics.

The weighted combination of the above feature maps forms the final global feature map, i.e. $\mathcal{F} = \alpha\ \mathcal{GM} + \beta\ \mathcal{SM} + \gamma\ \mathcal{DM}$. Figure 6 illustrates the importance of using scribbles and the proposed combination of energy maps. It is important to note that unlike some of the existing seam-based approaches [42], we generate the feature map only once for the input image.

3.2.3 Scribble-conditioned Seam Generation

For each scribble s, the paired end-points of the scribble and its corresponding upper pseudo-scribble u are connected to obtain an enclosed upper region U - see the block 'Region Masks' which is part of 'Scribble-Conditioned Seam-Generation' (shaded blue) in Fig. 4. The region's mask is applied to global feature map $\mathcal{F}$ and cropped to obtain the upper region feature map $\mathcal{F}_U$ for the scribble. To constrain the seams to lie within the masked portion, feature map values outside the mask are set to a fixed 'high energy' value. The upper region feature map is used during seam generation [5].

Algorithm 1 Scribble-Conditioned Text Line Polygon Generation (Sec. 3.2.3)

```
▷ Input binaryImage B and set of scribbles $\mathcal{S}$ from Stage I (Sec. 3.1)
▷ Output Set of text line polygons $\mathcal{P}$
$\theta \leftarrow$ ComputeGap($\mathcal{S}$)    ▷ Obtain interline gap using inter-scribble gap statistics
▷ Feature Map Generation
$\mathcal{GM} \leftarrow$ GenerateGradientMap($B, \mathcal{S}$)
$\mathcal{SM} \leftarrow$ GenerateSmoothingMap($B, \mathcal{S}$)
$\mathcal{DM} \leftarrow$ GenerateDiacriticMap($B, \mathcal{S}$)
$\mathcal{F} \leftarrow$ GenerateGlobalFeatureMap($\mathcal{GM}, \mathcal{SM}, \mathcal{DM}$)
▷ Scribble-conditioned Seam Generation
for $s$ in $\mathcal{S}$ do                                          ▷ For each scribble
    $u, l \leftarrow$GeneratePseudoScribbles($s, \theta$)
    ▷ Generate upper seam
    $U \leftarrow$ GetRegion($s, u$)
    $\mathcal{F}_U \leftarrow$ GetCroppedFeatureMap($U, \mathcal{F}$)
    $S_U \leftarrow$ GenerateSeams($\mathcal{F}_U$)
    ▷ Generate lower seam
    $L \leftarrow$ GetRegion($s, l$)
    $\mathcal{F}_L \leftarrow$ GetCroppedFeatureMap($L, \mathcal{F}$)
    $S_L \leftarrow$ GenerateSeams($\mathcal{F}_L$)
    ▷ Generate the final text line polygon
    $P \leftarrow$ GenerateLinePolygon($S_U, S_L$)
    $\mathcal{P} \leftarrow \mathcal{P} \cup \{P\}$
end for
return $\mathcal{P}$
```

For a $M \times N$ image, a horizontal seam R is a connected sequence of pixels and can be defined as $R = (x_i, y_i); i = 1, 2, \ldots r, 1 \leqslant x_i \leqslant N, 1 \leqslant y_i \leqslant M$ where $x_1 = 1, x_r = N$ and $|x_i - x_{i-1}| \leqslant 1, i = 2, 3, \ldots r$. The 'energy cost' of the seam is defined as $U(R) = \sum_{i=1}^{r} \mathcal{F}_U(x_i, y_i)$. The seam with the minimum cost is defined as $S_U = \arg\min_{R} U(R)$ and is found using dynamic programming. In this context, feature map $\mathcal{F}_U$ has been constructed such that the minimum energy seam corresponds to tight upper boundary of the associated text line. Additionally, to enhance the tight-fit characteristic of the seam, we induce a bias in choosing the lowest energy path. During the seam propagation step, we greedily pick the lowest x or y coordinate value among potential energy paths. This choice results in energy seams circumscribing the character components tightly. A similar procedure as above is repeated with the lower pseudo-scribble l to obtain a tight lower boundary seam S_L for the text line. These seams (S_U, S_L) are connected at their paired endpoints to obtain the final high precision polygon P enclosing the text line.

It is important to note that the scribble generated in Stage-I determines the sub-image region in which seam generation operates. Confining seam generation by using scribble-based masks helps produce compact enclosing boundaries (see Figure 6). This is unlike other seam-based methods which generate seams that go beyond actual extent of the text line. Algorithm 1 outlines the procedure for scribble-conditioned text line polygon generation.

4 Experiments

4.1 Datasets

We have tested the models on a selection of palm leaf manuscript datasets - Indiscapes2 [44], the datasets provided for the Challenge B (Text Line Segmentation) of the ICFHR 2018 Competition On Document Image Analysis Tasks for Southeast Asian Palm Leaf Manuscripts [24] containing manuscripts from Balinese, Khmer and Sundanese languages. In addition, a new manuscript collection called KgathaM has also been introduced.

Indiscapes2[38]: This is the largest dataset for Indic palm leaf manuscripts and consists of manuscripts sourced from four distinct sources. Indiscapes2 comprises of 1275 documents with a large diversity in scripts,language,semantic regions, document dimensions, number of lines and text line density. It has 748 manuscript leaves for training and 258 leaves for the test split. The average manuscript dimension is 750 × 1900.

KgathaM: We introduce this new collection of palm leaf manuscript written in a classical component-based script of the Indic language Malayalam. The manuscript contains verses from a poem. A unique aspect is that the poem is written on manuscript leaves continuously and end to end, without spaces between words. It has a total of 392 pages with $8-12$ lines in each document. We have considered 313 pages for train split and 79 pages in the test split. The manuscript leaves are quite dense with an average of 9-10 text lines and contain extremely small character components. The average size of the manuscript page is 400 × 2800.

Balinese[23]: This consists of Balinese manuscripts. It has been extracted from the AMADI LontarSet [23], with 393 pages of palm leaf manuscripts from 23 different collections. In general, the documents have 4 text lines, most of them double-columned with occasional illustrations. One common characteristic of this manuscript is the variety of diacritics. The Challenge provides a total of 96 pages with 47 pages in the train split and 49 pages in the test split. In general the pages have 4 text lines. The average size of the manuscript page is 500 × 5000.

Khmer[50]: This set consists of Khmer (Cambodian) manuscripts. It has been extracted from the SleukRith Set [50], with 657 pages of Khmer palm leaf manuscript randomly selected from different sources. In general, the pages have 5 text lines. The Challenge provides a total of 250 pages with 50 pages in the train split and 200 pages in the test split. The average size of the manuscript page is 500 × 5500.

Sundanese[48]: This set consists of Sundanese manuscripts. It has been extracted from the Sunda Set [48], with 66 pages of Sundanese Lontar randomly selected from 27 collections. The Challenge provides a total of 61 pages with 31 pages in the train split and 30 pages in the test split. On average, the pages consist of 4 text lines. The mean size of the manuscript page is 350 × 3000.

Table 1. Comparison of SeamFormer with existing approaches on benchmark datasets (Sec. 5).

	Indiscapes2[38]	KGathaM	Bali[23]	Sunda[48]	Khmer[50]
IoU ↑					
MMRCNN [38]	0.55	0.34	0.23	0.28	0.28
Palmira [44]	0.76	0.69	0.42	0.68	0.45
Doc-UFCN [9]	0.16	0.12	0.08	0.23	0.10
dhSegment [36]	0.34	0.12	0.03	0.12	0.08
LCG [2]	0.37	0.20	0.12	0.12	0.18
DocExtractor [30]	0.10	0.17	0.01	0.02	0.04
SeamFormer	**0.78**	**0.84**	**0.66**	**0.77**	**0.69**
HD ↓					
MMRCNN [38]	447.58	855.76	2106.30	1147.30	1760.48
Palmira [44]	73.32	57.84	1699.58	130.34	1190.95
Doc-UFCN [9]	339.30	238.87	1873.00	630.79	2552.26
dhSegment [36]	295.58	216.79	2232.90	394.16	1560.45
LCG [2]	207.76	346.93	797.51	367.60	496.31
DocExtractor [30]	806.17	1423.26	3552.19	1865.25	3987.37
SeamFormer	**21.91**	**16.05**	**48.86**	**32.18**	**48.37**
AvgHD ↓					
MMRCNN [38]	57.13	132.50	302.59	145.07	270.47
Palmira [44]	7.29	2.74	224.79	6.50	203.59
Doc-UFCN [9]	70.04	49.16	319.06	98.55	481.19
dhSegment [36]	60.33	43.60	415.24	66.77	319.57
LCG [2]	16.82	29.72	95.18	39.65	44.50
DocExtractor [30]	149.29	219.68	778.60	331.00	898.16
SeamFormer	**0.65**	**0.25**	**2.53**	**1.01**	**2.39**
HD95 ↓					
MMRCNN [38]	355.74	702.45	1766.12	918.85	1449.68
Palmira [44]	42.47	21.49	1393.15	49.06	1019.12
Doc-UFCN [9]	304.73	214.35	1628.43	520.38	2271.27
dhSegment [36]	262.83	192.33	1967.72	329.84	1380.09
LCG [2]	99.77	197.94	390.21	231.38	191.88
DocExtractor [30]	595.61	1084.01	3255.44	1656.47	3654.05
SeamFormer	**4.59**	**1.96**	**19.49**	**7.77**	**18.83**

4.2 Implementation Details

Stage-I: For the ViT network, we use 256×256 overlapping manuscript patches with appropriate padding. Resampling is used to overcome the imbalance between text and empty (non-text) patches. For training the binarizer branch for South-East Asian datasets, we use the binary dataset from Challenge A of the ICFHR 2018 contest [24]. For other datasets, we use Sauvola-Niblack binarisation [43,33] as the ground truth. We initialize the binarization branch with

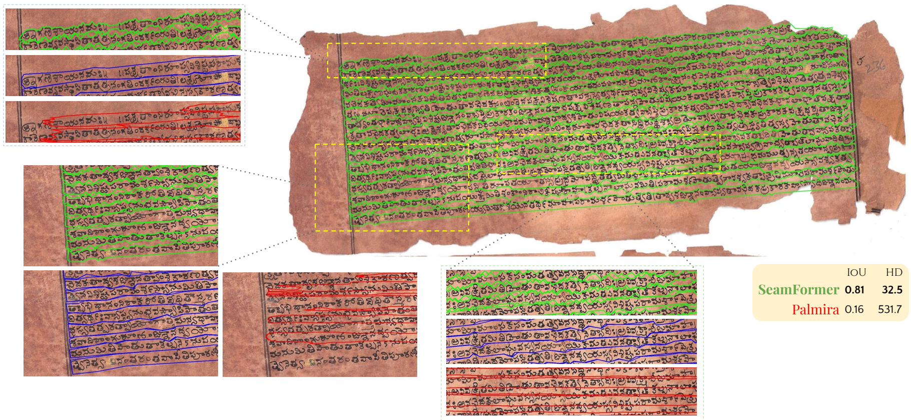

Figure 7. A challenging manuscript from Indiscapes2 [44]. The figure shows insets of regions with ground-truth (blue) and predictions from SeamFormer (green) and Palmira [44] (red). The document level performance scores are shown in bottom right.

pre-trained weights [47]. The learning rate is initialized to 0.05 and is decayed by Pytorch's learning scheduler, ExponentialLR with $\gamma = 0.8$. For training both of these branches we leverage the L2 loss. We adopt a training procedure where every individual branch is trained separately, while the other branch's weights are frozen. The optimizer used is stochastic gradient descent with $\gamma = 0.1$ and momentum of 0.9. We perform data-parallel optimization distributed across 2 GeForce RTX 2080 Ti GPUs for 40 epochs, with a fixed batch size of 4. We use random rotation augmentation $\alpha \in (-30, 30)$ to improve performance for non-axis oriented manuscripts. To tackle varied manuscript background textures and noise, we apply Gaussian Noise, AdvancedBlur, RandomColor, RandomFog, RandomBrightness and HueSaturations augmentations [11]. For post-processing, we apply erosion filters - a horizontal rectangular kernel 1×11 thrice, followed by a 1×1 dilation to separate any overlapping medial blobs. These blobs undergo a skeletonization procedure followed by pruning to remove any spurious branches with a minimum area threshold of 100 pixels. The post-processing is robust and does not need to be changed across datasets or approaches.

Stage-II: The offset for pseudo-scribble generation δ is set to 5. In the feature map generation pipeline, we use the standard 3×3 Sobel kernel for Gradient Map. We apply a Gaussian blur kernel of 15×11 for high spatial coverage within the image to compute the Smoothing Map. The weights for various feature maps are empirically set to $\alpha = 0.4$ ($\mathcal{GM}$), $\beta = 0.6$ ($\mathcal{SM}$) and $\gamma = 1.0$ ($\mathcal{DM}$). The global feature map is normalised to $[0, 1]$ before the seam generation process.

Table 2. Ablation experiments using Indiscapes2. Proposed refers to design choices in SeamFormer.

Row-id	Stage I	Stage II	IoU ↑	HD ↓	HD_{95} ↓	Avg.HD ↓
1	Text Baseline	Proposed	0.63	62.59	8.29	35.86
2	ARU-Net [18]	Proposed	0.69	103.57	9.37	51.51
3	Proposed	$\mathcal{GM}$	0.76	23.83	0.78	5.15
4	Proposed	$\mathcal{GM}, \mathcal{SM}$	0.77	22.40	0.71	4.71
5	**Proposed**	**Proposed**	**0.78**	**21.91**	**0.65**	**4.59**

5 Results

For quantitative evaluation of text line segmentation, we compare SeamFormer against various state-of-the-art approaches developed for palm-leaf and other types of manuscripts. The approaches were fine-tuned for each dataset. As performance measure, we use IoU. In their work, Trivedi et al. [49] show that Hausdorff Distance (HD) and its variants - HD_{95} and Average HD reflect the prediction performance for polygon boundary predictions better than area-based IoU metric. Therefore, we report these measures as well. Note that smaller the HD-based scores, better the text line polygon prediction.

The overall quantitative results can be seen in Table 1. SeamFormer clearly outperforms the competing strong baseline approaches across all the datasets and across the performance measures. This shows the generalizability provided by our approach. Our consistently small HD scores are due to the high precision polygons generated by our custom scribble-conditioned seam generation pipeline. For some existing approaches, HD-based scores are one or two orders of magnitude higher due to low line accuracy. Most of these approaches resize the input image to a fixed size for optimal training of the neural network. However, due to the extremely large aspect ratio ($\approx 10 : 1$) and range in sizes for palm leaf manuscripts, the resizing causes text line polygon aliasing, causing poor performance. These factors are not an issue for SeamFormer since resizing is not a part of the pipeline. The second-best network Palmira [44] is competitive in terms of IoU for Indiscapes2 [44]. However, the performance gap is substantial for other datasets and other performance measures as well.

5.1 Ablation Study

We perform an ablation analysis with Indiscapes2 dataset to determine the contribution of various design choices within Stage-I (scribble generation) and Stage-II (seam generation). Instead of a medial scribble through the text, we tried the popular text underline (baseline) as an alternative. The bottom of the text line polygon is used as the baseline. However, this led to sub par performance since the baseline is not guaranteed to touch the text and does not prevent seams

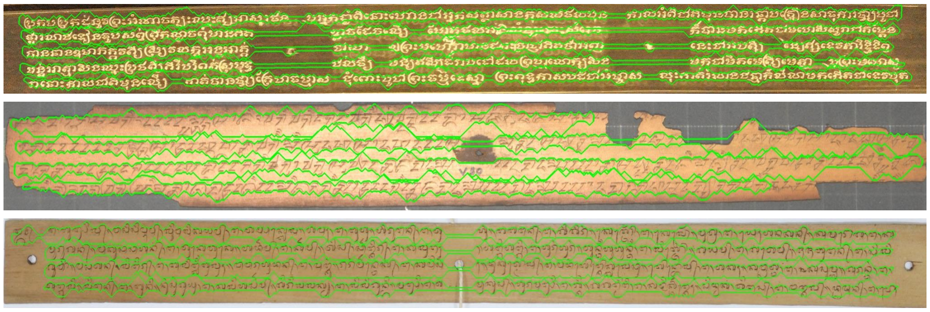

Figure 8. SeamFormer predictions on South-East Asian Manuscripts [24] – Khmer (top), Sundanese (middle) and Balinese (top).

from cutting in between text components of the lines (row 1 of Table 2). In another experiment, we re-trained the popular ARU-Net [18] as an alternative to our ViT architecture for obtaining scribbles. ARU-Net produces disconnected scribbles which results in poor performance (row 2). Keeping Stage-I fixed, we also conducted experiments to determine the impact of each feature map (rows 3-4). We observe that the full set of feature maps (last row) provides the best performance – also see Figure 6.

5.2 Qualitative Results

A visual comparison of performance between ground-truth and predictions by SeamFormer and the second-best model Palmira [44] can be seen in Figure 7. The effect of resizing can be seen in Palmira's incorrect and coarse predictions. Despite the challenging nature of the manuscript (e.g. document tilt, dense and unevenly spaced text lines), SeamFormer predictions are significantly more accurate. This trend can also be seen in sample manuscripts from other datasets - see Figures 8,9,10.

6 Conclusion

We introduce SeamFormer, a novel approach for high precision text line segmentation in handwritten documents. Instead of a monolithic framework, we tackle the challenge of text line segmentation using a divide-and-conquer two stage approach. The first stage generates medial line 'scribbles' which provide crucial information about the curvature of the text line and a binarized version of the input image. In the second stage, these scribbles and custom-designed feature maps derived from the binarized image are fed to a seam generation algorithm which generates the desired tight-fitting line polygons.

Our approach is a resizing-free method. As a result, text line gaps are not distorted or aliased, leading to significantly better results. Our novel inclusion of Diacritic Map in the second stage ensures complete and correct inclusion

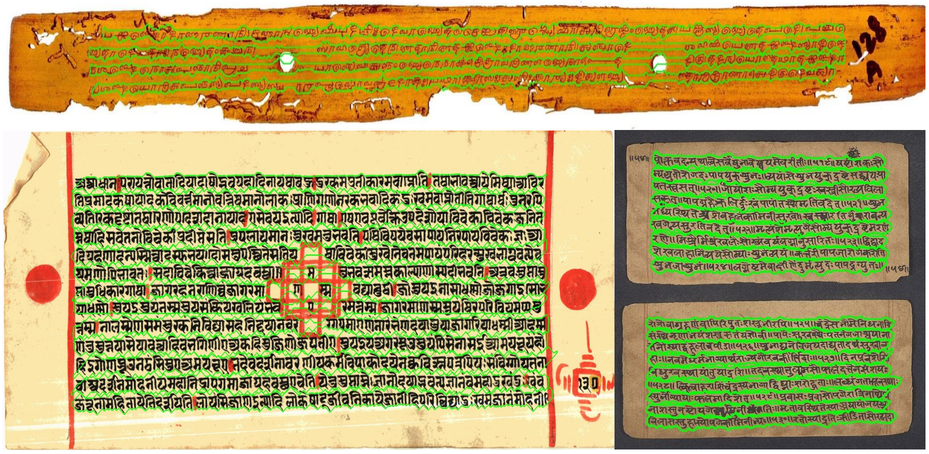

Figure 9. SeamFormer predictions on Indiscapes2 [44] manuscripts.

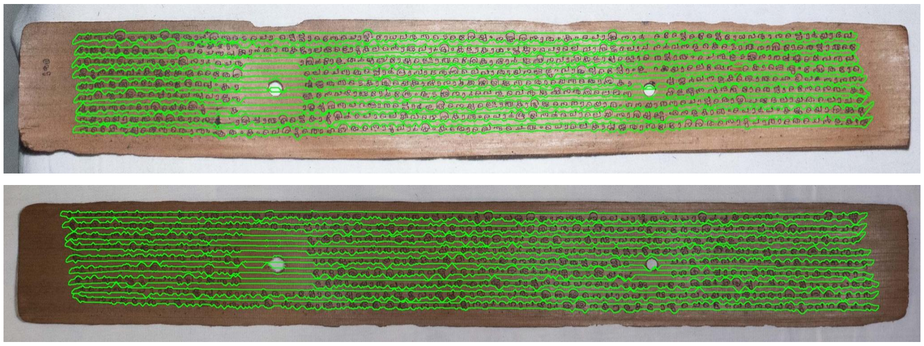

Figure 10. SeamFormer predictions on manuscripts from the newly introduced KGathaM collection.

of diacritics within the predicted polygon. Also, pseudo-scribbles are a key innovation in our approach. The pseudo-scribbles serve as energy barriers during seam generation and ensure the seams do not cross the text line's spatial extents. The pseudo-scribbles also prevent the seams from deviating too much from the reference line unlike some existing approaches. The efficacy of our approach is evident from its comparatively superior performance across challenging datasets and metrics.

An additional advantage of our approach is that it enables interactive human-in-the-loop refinement. For instance, scribbles could be manually added for any missed lines followed by second stage processing. Another advantage is that unlike some existing approaches, no post-processing on the polygons is required. Our results demonstrate the utility of SeamFormer for line segmentation across multiple challenging datasets. Overall, SeamFormer is an attractive option for generating precise text line polygons in handwritten manuscript collections. The source code, pretrained models and associated material are available at this link: https://ihdia.iiit.ac.in/seamformer.

References

1. Alaei, A., Pal, U., Nagabhushan, P.: A new scheme for unconstrained handwritten text-line segmentation. Pattern Recognition **44**(4), 917–928 (2011) 3
2. Alberti, M., Vögtlin, L., Pondenkandath, V., Seuret, M., Ingold, R., Liwicki, M.: Labeling, cutting, grouping: an efficient text line segmentation method for medieval manuscripts. In: 2019 International Conference on Document Analysis and Recognition (ICDAR). pp. 1200–1206. IEEE (2019) 4, 11
3. Arvanitopoulos, N., Süsstrunk, S.: Seam carving for text line extraction on color and grayscale historical manuscripts. In: 2014 14th International Conference on Frontiers in Handwriting Recognition. pp. 726–731. IEEE (2014) 4
4. Asi, A., Saabni, R., El-Sana, J.: Text line segmentation for gray scale historical document images. In: Proceedings of the 2011 workshop on historical document imaging and processing. pp. 120–126 (2011) 4
5. Avidan, S., Shamir, A.: Seam carving for content-aware image resizing. In: ACM SIGGRAPH 2007 papers, pp. 10–es (2007) 4, 8
6. 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). pp. 374–379. IEEE (2018) 3
7. 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). pp. 2304–2311. IEEE (2021) 3
8. Barakat, B.K., El-Sana, J., Rabaev, I.: The pinkas dataset. In: 2019 International Conference on Document Analysis and Recognition (ICDAR). pp. 732–737. IEEE (2019) 1
9. Boillet, M., Kermorvant, C., Paquet, T.: Multiple document datasets pre-training improves text line detection with deep neural networks. In: 2020 25th International Conference on Pattern Recognition (ICPR). pp. 2134–2141. IEEE (2021) 3, 11
10. Bruzzone, E., Coffetti, M.C.: An algorithm for extracting cursive text lines. In: Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR'99 (Cat. No. PR00318). pp. 749–752. IEEE (1999) 3
11. Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M., Kalinin, A.A.: Albumentations: Fast and flexible image augmentations. Information **11**(2) (2020). https://doi.org/10.3390/info11020125, https://www.mdpi.com/2078-2489/11/2/125 12
12. Chamchong, R., Fung, C.C.: Character segmentation from ancient palm leaf manuscripts in thailand. In: Proceedings of the 2011 Workshop on Historical Document Imaging and Processing. pp. 140–145 (2011) 3
13. Chamchong, R., Fung, C.C.: Text line extraction using adaptive partial projection for palm leaf manuscripts from thailand. In: 2012 International Conference on Frontiers in Handwriting Recognition. pp. 588–593. IEEE (2012) 3

14. Clausner, C., Antonacopoulos, A., Derrick, T., Pletschacher, S.: Icdar2019 competition on recognition of early indian printed documents–reid2019. In: 2019 International Conference on Document Analysis and Recognition (ICDAR). pp. 1527–1532. IEEE (2019) 1
15. Dolfing, H.J., Bellegarda, J., Chorowski, J., Marxer, R., Laurent, A.: The "scribblelens" dutch historical handwriting corpus. In: 2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR). pp. 67–72. IEEE (2020) 1
16. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. ICLR (2021) 5
17. Grüning, T., Labahn, R., Diem, M., Kleber, F., Fiel, S.: Read-bad: A new dataset and evaluation scheme for baseline detection in archival documents. In: 2018 13th IAPR International Workshop on Document Analysis Systems (DAS). pp. 351–356. IEEE (2018) 1
18. Grüning, T., Leifert, G., Strauß, T., Michael, J., Labahn, R.: A two-stage method for text line detection in historical documents. International Journal on Document Analysis and Recognition (IJDAR) **22**(3), 285–302 (2019) 13, 14
19. He, J., Downton, A.C.: User-assisted archive document image analysis for digital library construction. In: Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings. pp. 498–502. IEEE (2003) 3
20. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: ICCV (2017) 4
21. Jindal, A., Ghosh, R.: Text line segmentation in indian ancient handwritten documents using faster r-cnn. Multimedia Tools and Applications pp. 1–20 (2022) 3
22. Kesiman, M.W.A., Burie, J.C., Ogier, J.M.: A new scheme for text line and character segmentation from gray scale images of palm leaf manuscript. In: 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR). pp. 325–330. IEEE (2016) 3
23. Kesiman, M.W.A., Burie, J.C., Wibawantara, G.N.M.A., Sunarya, I.M.G., Ogier, J.M.: Amadi_lontarset: the first handwritten balinese palm leaf manuscripts dataset. In: 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR). pp. 168–173. IEEE (2016) 10, 11
24. Kesiman, M.W.A., Valy, D., Burie, J.C., Paulus, E., Suryani, M., Hadi, S., Verleysen, M., Chhun, S., Ogier, J.M.: 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). pp. 483–488 (2018). https://doi.org/10.1109/ICFHR-2018.2018.00090 10, 11, 14
25. Kesiman, M.W.A., et al.: Benchmarking of document image analysis tasks for palm leaf manuscripts from southeast asia. Journal of Imaging **4**(2), 43 (2018) 2, 3
26. Kleber, F., Fiel, S., Diem, M., Sablatnig, R.: Cvl-database: An off-line database for writer retrieval, writer identification and word spotting. In: 2013 12th international conference on document analysis and recognition. pp. 560–564. IEEE (2013) 1
27. Li, D., Wu, Y., Zhou, Y.: Linecounter: Learning handwritten text line segmentation by counting. In: 2021 IEEE International Conference on Image Processing (ICIP). pp. 929–933. IEEE (2021) 3, 4
28. Likforman-Sulem, L., Faure, C.: Extracting text lines in handwritten documents by perceptual grouping. Advances in handwriting and drawing: a multidisciplinary approach pp. 117–135 (1994) 3

29. Mechi, O., Mehri, M., Ingold, R., Amara, N.E.B.: Text line segmentation in historical document images using an adaptive u-net architecture. In: 2019 International Conference on Document Analysis and Recognition (ICDAR). pp. 369–374. IEEE (2019) 3
30. Monnier, T., Aubry, M.: docExtractor: An off-the-shelf historical document element extraction. In: ICFHR (2020) 3, 11
31. Nagy, G., Seth, S.C., Stoddard, S.D.: Document analysis with an expert system. In: Pattern recognition in practice II. pp. 149–155 (1985) 3
32. Nguyen, T.N., Burie, J.C., Le, T.L., Schweyer, A.V.: An effective method for text line segmentation in historical document images. In: 2022 26th International Conference on Pattern Recognition (ICPR). pp. 1593–1599. IEEE (2022) 4
33. Niblack, W.: An introduction to digital image processing. Strandberg Publishing Company (1985) 11
34. Nikolaidou, K., Seuret, M., Mokayed, H., Liwicki, M.: A survey of historical document image datasets. Int. J. Doc. Anal. Recognit. **25**(4), 305–338 (dec 2022). https://doi.org/10.1007/s10032-022-00405-8, https://doi.org/10.1007/s10032-022-00405-8 3
35. O'Gorman, L.: The document spectrum for page layout analysis. IEEE Transactions on pattern analysis and machine intelligence **15**(11), 1162–1173 (1993) 3
36. Oliveira, S.A., Seguin, B., Kaplan, F.: dhsegment: A generic deep-learning approach for document segmentation. In: 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR). pp. 7–12. IEEE (2018) 3, 11
37. Pavildas, T.: Page segmentation by white streams. In: Proc. 1st Int. Conf. Document Analysis and Recognition. pp. 945–953 (1991) 3
38. Prusty, A., Aitha, S., Trivedi, A., Sarvadevabhatla, R.K.: Indiscapes: Instance segmentation networks for layout parsing of historical indic manuscripts. In: ICDAR. pp. 999–1006 (2019) 2, 4, 10, 11
39. Pu, Y., Shi, Z.: A natural learning algorithm based on hough transform for text lines extraction in handwritten documents. In: Advances in Handwriting Recognition, pp. 141–150. World Scientific (1999) 3
40. Renton, G., Soullard, Y., Chatelain, C., Adam, S., Kermorvant, C., Paquet, T.: Fully convolutional network with dilated convolutions for handwritten text line segmentation. International Journal on Document Analysis and Recognition (IJDAR) **21**, 177–186 (2018) 3
41. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015) 3
42. Saabni, R., El-Sana, J.: Language-independent text lines extraction using seam carving. In: 2011 International Conference on Document Analysis and Recognition. pp. 563–568. IEEE (2011) 4, 8
43. Sauvola, J., Pietikäinen, M.: Adaptive document image binarization. Pattern recognition **33**(2), 225–236 (2000) 11
44. Sharan, S., Aitha, S., Kumar, A., Trivedi, A., Augustine, A., Sarvadevabhatla, R.K.: Palmira: a deep deformable network for instance segmentation of dense and uneven layouts in handwritten manuscripts. In: International Conference on Document Analysis and Recognition. pp. 477–491. Springer (2021) 2, 4, 10, 11, 12, 13, 14, 15
45. Shi, Z., Setlur, S., Govindaraju, V.: Text extraction from gray scale historical document images using adaptive local connectivity map. In: Eighth International Conference on Document Analysis and Recognition (ICDAR'05). pp. 794–798. IEEE (2005) 3

46. Shi, Z., Setlur, S., Govindaraju, V.: A steerable directional local profile technique for extraction of handwritten arabic text lines. In: 2009 10th International Conference on Document Analysis and Recognition. pp. 176–180. IEEE (2009) 3
47. Souibgui, M.A., Biswas, S., Jemni, S.K., Kessentini, Y., Fornés, A., Lladós, J., Pal, U.: Docentr: An end-to-end document image enhancement transformer. In: 2022 26th International Conference on Pattern Recognition (ICPR) (2022) 12
48. Suryani, M., Paulus, E., Hadi, S., Darsa, U.A., Burie, J.C.: The handwritten sundanese palm leaf manuscript dataset from 15th century. In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). vol. 1, pp. 796–800. IEEE (2017) 10, 11
49. Trivedi, A., Sarvadevabhatla, R.K.: Boundarynet: An attentive deep network with fast marching distance maps for semi-automatic layout annotation. In: International Conference on Document Analysis Recognition, ICDAR 2021 (2021) 13
50. Valy, D., Verleysen, M., Chhun, S., Burie, J.C.: A new khmer palm leaf manuscript dataset for document analysis and recognition: Sleukrith set. In: Proceedings of the 4th International Workshop on Historical Document Imaging and Processing. pp. 1–6 (2017) 10, 11
51. Valy, D., Verleysen, M., Sok, K.: Line segmentation for grayscale text images of khmer palm leaf manuscripts. In: 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA). pp. 1–6. IEEE (2017) 3
52. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017) 5
53. Yalniz, I.Z., Manmatha, R.: A fast alignment scheme for automatic ocr evaluation of books. In: 2011 International Conference on Document Analysis and Recognition. pp. 754–758. IEEE (2011) 1
54. Zahour, A., Taconet, B., Mercy, P., Ramdane, S.: Arabic hand-written text-line extraction. In: Proceedings of Sixth International Conference on Document Analysis and Recognition. pp. 281–285. IEEE (2001) 3