Good morning everyone,

Before we begin, let me introduce my self-briefly. I am Mechi Olfa and I am a PhD student at laboratory of Advanced Technology and Intelligent Systems-ENISo-Tunisia.

Today, I am going to present my ICPR poster about combining deep and ad-hoc solutions to localize text lines in ancient Arabic document images.

During the last 30 years, numerous libraries and archives have conducted large digitization programs of valuable cultural heritage of ancient documents. In fact, ground need for efficient multilingual handwritten text transcription tools has been recently flagged by digital humanities and archives in order to assist experts in manual transcription provide an efficient multilingual text indexing and retrieval of archival document images and to access to large sets of cultural heritage documents.

In this context, our research work has been carried out as part of a big research project with the Tunisian Ministry of Higher Education and Scientific Research and the collaboration of the Tunisian national archives (ANT) and our partner industry “Smart Information trade”.

Our research project aims at designing an automatic system able to transcribe and index automatically digitized archival documents. However, to ensure an efficient text transcription from archival handwritten and printed image, a robust text line segmentation task is required.

In this context, our work tackles the task of localizing text lines in ancient Arabic documents by proposing a hybrid method combining a deep network with ad-hoc document image analysis techniques.

Figure 1 illustrates the proposed method, which is conceptualized by the three following modular processes. The first step consists on extraction of the X-height contours using U-Net architecture. The second step is the extraction of the foreground contour by applying modified RLSA algorithm. Afterward, we compute the intersection area between each foreground contour from binary image and each x-height contours. There are 3 cases.

-Number of intersection areas= 0 That means that the analyzed foreground contour does not necessarily correspond to a textual content (noise or graphic).

-Number of intersection areas= 1 That means that the analyzed foreground contour belongs to exactly a single text line.

-Number of intersection areas>1 that means that the analyzed foreground contour belongs to more than a single text line. Hence, it is appropriate to split the extracted foreground contour into smaller

ones. The split process is firstly based on assigning to the foreground contour, the X-height (Xhm) having the largest intersection area with the foreground contour. Then, the distances between each point belonging to the foreground contour and the medians of all predicted X-heights are computed in order to retain only the foreground contour points that have minimal distance to the Xhm. Indeed, the retained foreground contour points define a novel foreground contour belonging to a single text lines. The split process is applied recursively until none foregrounds contour point remains non-assigned to a single X-height. Since each obtained foreground contour has been assigned to a single X-height, the ascender and descender components of each text line can be finally determined. The ascender and descender components of each text line correspond to the highest and lowest points of the different obtained foreground contours belonging to a single X-height, respectively.

Figure 2 illustrates examples of historical document images of the four evaluated datasets: RASM, ANT-Arabic, cBAD and ANT-Latin.

In our experimental, we have computed the same per-pixel accuracy metrics used in the context of the cBAD 2017 competition: precision (P), recall (R) and F-measure (F). Furthermore, five performance evaluation metrics have been computed: “Match”, “Miss”, “False alarm”, “Split” and “Merge” to evaluate the performance of our proposed approach.

Table 1 and 2 illustrates the performance evaluation results of the proposed method on ancient Arabic and Latin document images.

By comparing the performance of the proposed method with those obtained with the five methods presented in the RASM 2018 contest (Tesseract 3, Tesseract 4, FRE11, KFCN and RDI), we observe that the proposed method achieves the best performance as illustrated in table II.

Figures 3(a) and 3(b) depict the obtained results obtained by using the proposed method on document images.

However, the proposed method has some limitations. Some text lines are not correctly detected due to the biased output of the U-Net stage. On the other side, some text lines include parts of their adjacent text lines. Nevertheless, it has been shown that these limitations will not affect the performance of the text recognition task.

I would like to end today’s presentation by summarizing the main information of our poster presentation.

In our work, we combine deep and ad-hoc solutions to localize text lines in ancient Arabic document images. The proposed method achieved a satisfactory result for both Arabic and Latin

document images. There are several possibilities that stem from this work such as evaluate the proposed method on other public ancient handwritten Arabic document image datasets (e.g., Hadara and VML-HD) on one hand, and propose a deep system able to transcribe printed and handwritten text lines from Arabic and Latin historical document images on the other hand.

For any question, please feel free to contact the first author by email.