eScriptorium comparison

Berat Kurar-Barakat and Nachum Dershowitz

12.09.2022

eScriptorium is an open source document image analysis platform with a web interface that allows users to segment and transcribe document images [6]. The segmentation module can perform region segmentation and text line segmentation. Region segmentation refers to finding the locations of textual and non-textual zones. Text line segmentation refers to finding the locations of individual text lines. Furthermore, the platform can enrich the segmented units with semantic information that is their classification as title, paragraph, ornament, illustration, etc. [5].

Their legacy segmenter is a learning free algorithm but requires binarization and manhattan layout [6]. Our learning free text line segmentation method [7] also requires binarization however, we modified it to be able segment complex layout [1]. A performance comparison is not possible as they did not report results with the learning free method.

They also developed a learning based algorithm that is able to segment non-binary and complex layout documents [5]. The learning based method has three stages; pixel classification, baseline extraction and polygonization, and region extraction. Pixel classification uses a neural network that classifies the pixels as baselines that strikethrough the text lines, regions that cover the textual zones, and auxiliary classes that shows the beginning and the end of the text lines. The second stage extracts polylines from the baseline and auxiliary pixels using thresholding and skeletonization, then extracts text line bounding polygons from the polylines using seam-carving. The final stage extracts region bounding polygons from the region pixels using thresholding and contour finding algorithm. All the three stages of learning-based method is layout agnostic.

Our learning based methods for region and text line segmentation are not holistic as theirs. Our region segmentation method [4] also uses a neural network to segment and semantically classify complex layout regions. A performance comparison for region segmentation is not possible as they did not report results.

Our learning based text line segmentation method [2] also uses a neural network to classify the pixels as baseline or background pixel. The baseline pixels are then thresholded to form blob lines that strikethrough the text lines. Then we use an energy minimization function that is guided by the blob lines for assigning the components to individual text lines. Hence, all our three stages are also layout agnostic. However our energy minimization stage requires binarization in contrast to their seam carving stage. A basic methodological difference is that they input the downscaled whole document image whereas we input the patches of document image for decreasing the memory requirement. Downscaling caused closely written textlines to merge hence their method is incapable of segmenting close text lines. On the other hand, the both, theirs and ours, suffer

from segmenting rare cases because the network did not see sufficient amount of training data on these cases. A performance comparison among the learning based text line segmentation methods is not possible as there are no results on a common dataset. However on cBAD dataset, our learning free method achieves an f-measure of 86.38 whereas their learning based method achieves an f-measure of 90.40.

Due to the problem of training data scarcity we proposed unsupervised learning based methods [3,8]. These methods differ only in the pixel classification stage where they do not require annotated data. However, they do not achieve a superior performance in comparison to the supervised methods albeit they are the first examples of an unsupervised learning based method for text line segmentation.

References

1. Barakat, B.K., 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). vol. 6, pp. 13–18. IEEE (2019)
2. 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: International Conference on Pattern Recognition. pp. 126–140. Springer (2021)
3. 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)
4. Barakat, B.K., El-Sana, J.: Binarization free layout analysis for arabic historical documents using fully convolutional networks. In: 2018 IEEE 2nd International Workshop on Arabic and Derived Script Analysis and Recognition (ASAR). pp. 151–155. IEEE (2018)
5. Kiessling, B.: A modular region and text line layout analysis system. In: 2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR). pp. 313–318. IEEE (2020)
6. Kiessling, B., Tissot, R., Stokes, P., Ezra, D.S.B.: escriptorium: An open source platform for historical document analysis. In: 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW). vol. 2, pp. 19–19. IEEE (2019)
7. Kurar Barakat, B., Cohen, R., Droby, A., Rabaev, I., El-Sana, J.: Learning-free text line segmentation for historical handwritten documents. Applied Sciences **10**(22), 8276 (2020)
8. Kurar Barakat, B., Droby, A., Saabni, R., El-Sana, J.: Unsupervised learning of text line segmentation by differentiating coarse patterns. In: International Conference on Document Analysis and Recognition. pp. 523–537. Springer (2021)