EasyChair Preprint

№ 9141

Utilizing 3D Information from Point Clouds to Support Document Image Binarization

Quang-Vinh Dang and Guee-Sang Lee

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

October 26, 2022

Utilizing 3D Information from Point Clouds to Support Document Image Binarization

Quang-Vinh Dang[1] and Guee-Sang Lee[1]

[1] Chonnam National University, Gwangju, Korea
quangvinh242003@yahoo.com

Abstract. One important step for the document processing domain is binarization for degraded document images. We introduce a new method that supports the network using 2D images. This support is based on new features of 3D point clouds converted from 2D images. Specifically, the original 2D images are fed to the LadderNet network to extract 2D information. Besides, the converted 3D point clouds from the original 2D images are fed to network for 3D point cloud input to get unique features. These features are useful and different from 2D information. So, fusing the output features from two of these architectures achieves a better result. We illustrate the benefit of this idea for binarizing documents.

Keywords: Image Binarization, Fusion network, 3D Point Clouds, Document

1 Introduction

For real situations where document images are substantially degraded due to poor storage methods and inappropriate maintenance conditions, document binarization is possible challenges. The important role of the binarization process has resulted in the group organizing the Document Image Binarization Competition (DIBCO)[1] in accordance with the 2013 ICDAR and ICFHR conferences[2]. A variety of methods for binarization have been developed with the advent of powerful deep convolutional neural network (CNN) models for object recognition, detection, and segmentation. In recent years, few deep learning methods [3, 4, 5, 6] have also been utilized for document images binarization. However, These deep learning models only make use of 2D images without 3D information that can be converted from original 2D images. In [7], the 2D image is converted into a 3D point cloud on a globe in which the distance between a point and the center of the globe is measured by the intensity. Authors discover new features that can help to denoise in degraded document images. However, this method is based on traditional method and doesn't use any useful 2D information. So, we propose a model that can take advantage of both 2D and 3D information.

The main idea of our algorithm is that pixels are transformed to points that have x, y, z in R^3 corresponding with the location and intensity of pixels in images as in [7]. These converted points are fed to the network for 3D point clouds. The output of the network has features containing useful information for document image binarization. Furthermore, the original 2D images are fed to another network for 2D images to get main features for final result. To integrate the helpful information from 2 of these different inputs, we apply the minimum function on the outputs from 2 of these architectures to combine them. The rest of the paper is organized as follows: Section 2 describes the proposed method for document image binarization. Section 3 reports our experiment results. Finally, the paper is concluded in section 4

2 Proposed Method

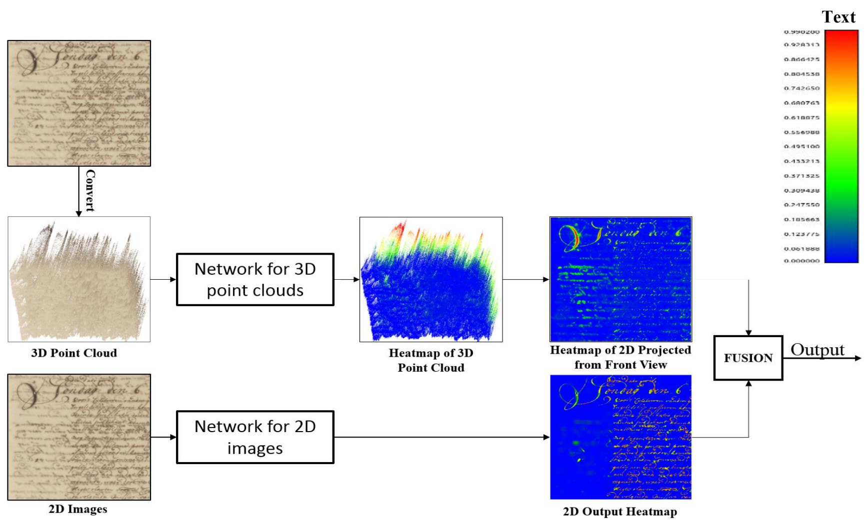


Fig. 1. Diagram of the proposed model. 3D point cloud inputs are derive from the pixel's (x, y) coordinates spreading the points over a 3D surface, and we transfom pixel intensity to z-component in three-dimensional Euclidean space R^3. 3D point clouds and 2D images are the inputs of network based on [8] and LadderNet[9], respectively. The outputs of each network are illustrated by heatmaps. The probability of text stroke has value corresponding with colour range in heatmap.The final output is fused by 2 thresholded outputs from each network.

The architecture of the proposed model is the fusion of the architecture [8] for 3D point clouds and the network LadderNet [9] for 2D images as in Fig. 1. In [8], the authors present the new design of Kernel Point Convolution (KPConv) that can apply for 3D point cloud input. In [9], the LadderNet is designed to have more encoder-decoder pairs that are effective in segmentation tasks.

The binary output from the architecture for 2D image input contains the main information of the final result. The other one has supporting information to the final result as in Fig. 2. We recognize the output with 3D point cloud input has low performance. However, it brings helpful and unique information that doesn't contain in the output with 2D image input. In Fig. 2(c), the orange area has a quite bad result but the remaining area has a relatively good result. In Fig. 2(d), it shows the completely opposite result. The final result has better performance by combining the advantages of both these outputs as in Fig. 2(e). Looking at Fig. 2 (d), the obtained part from 3D point clouds is a part of larger stroke while the obtained part from 2D images is a part of smaller stroke in Fig. 2 (c).

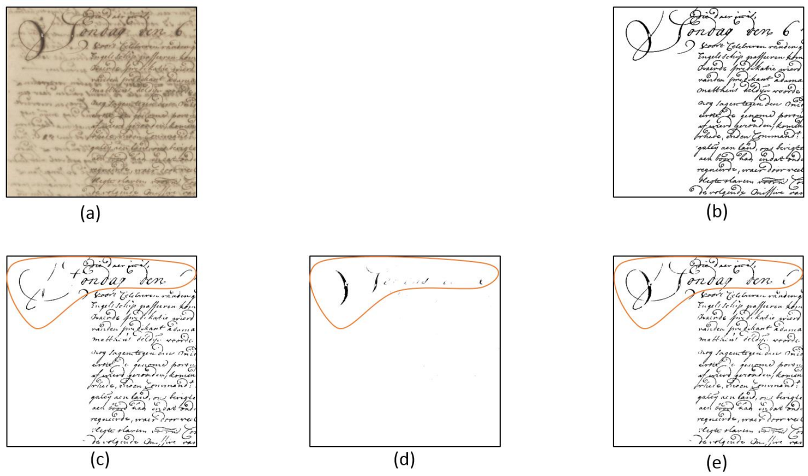


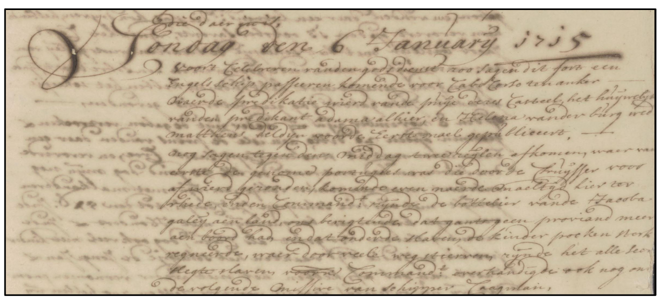
Fig. 2. Binarization Outputs. (a) Original image, (b) Ground-truth, (c) Output from the 2D network after thresholding the predicted map, (d) Output from the network for 3D point cloud inputs after thresholding the predicted map. It supports information for the remaining output as the orange area, (e) Fused output, created from a minimum of (c) and (d).

3 Experiment and Evaluation

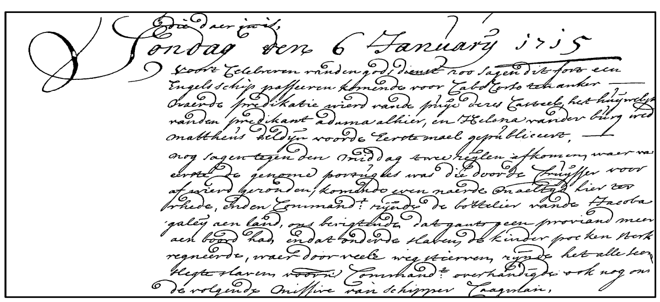
There are several public data sets for document binarization. Similar as [10] and [11], We select images on DIBCO 2009 [12], H-DIBCO 2010 [13], DIBCO 2011 [14], H-DIBCO 2012 [15], DIBCO 2013 [16], H-DIBCO 2014 [17], DIBCO 2016[18], for training set.The document images on DIBCO 2017 [10] is selected for the testing set. Evaluation metrics follow ICDAR Competition [10] such as F-measure (FM), pseudo-Fmeasure (F_{ps}), Peak signal-to-noise ratio (PSNR) and Distance Reciprocal Distortion Metric (DRD).

Table 1. Results of testing on DIBCO 2017 dataset

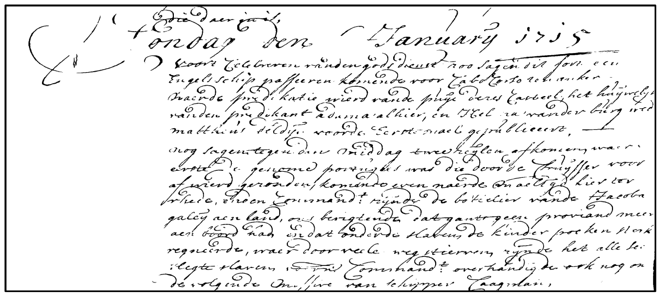
Methods	F-measure	Fps	PSNR	DRD
Otsu[19]	77.73	77.89	13.85	15.54
Niblack[20]	75.03	79.52	13.02	13.63
Winner[10]	91.04	92.86	18.28	3.40
Dang et al[11]	91.92	95.06	18.41	2.84
Network for 2D image	**90.01**	**93.40**	**17.74**	**3.59**
Ours	**91.33**	**93.08**	**17.99**	**3.31**



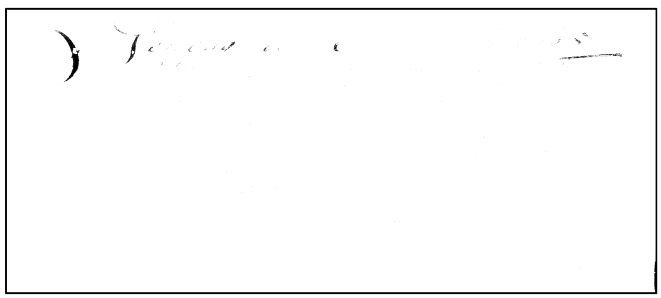
(a)



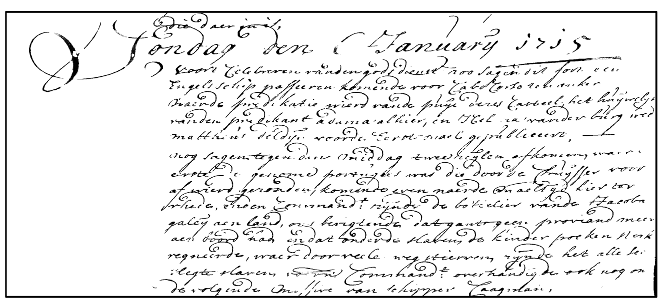
(b)



(c)



(d)



(e)

Fig. 3. Binarization results of the sample document image in the DIBCO 2017 dataset produced by different methods: (a) Original image, (b) The ground truth, (c) Output from the 2D network, (d) Output from 3D point cloud inputs, (e) Fused output.

4 Conclusion

A novel model is proposed in this paper. In document image binarization, integrating two distinct architectures for 2D image and 3D point cloud input has advantages. This is because we take advantages of each architecture's efficiency. We plan to develop a new framework for future work to improve results.

Acknowledgement

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF - 2018R1D1A3B05049058 & NRF – 2017R1A4A1015559). The corresponding author is Guee-Sang Lee

References

1. Lucas, S., Panaretos, A., Sosa, L., Tang, A., Wong, S., Young, R., Ashida, K., Nagai, H., Okamoto, M., Yamamoto, H., Miyao, H., Zhu, J., Ou, W., Wolf, C., Jolion, J., Todoran, L., Worring, M., Lin, X.: ICDAR 2003 Robust Reading Competitions: Entries, Results, and Future Directions. In: Int. J. Doc. Anal. Recognit. Vol. 7, pp. 105–122, (2005).
2. Stamatopoulos, N., Gatos, B., Louloudis, G., Pal, U., Alaei, A.: ICDAR 2013 Handwriting Segmentation Contest. In: 12th International Conference on Document Analysis and Recognition (ICDAR), pp. 1402-1406, (2013).
3. Tensmeyer, C., Martine, T.: Document Image Binarization with Fully Convolutional Neural networks. In: 14th IAPR International Conference on Document Analysis and Recognition, (2017).
4. Vo, Q., Kim, S., Yang, H., Lee, G.: Binarization Of Degraded Document Images Based On Hierarchical Deep Supervised Network. Journal Pattern Recognition, Volume 74 Issue C Pages 568-586, (2018).
5. Jorge, C., Gallego, A.: A Selectional Auto-encoder Approach for Document Image Binarization. Journal Pattern Recognition, Volume 86, (2017).
6. Peng, X., Cao, H., Natarajan, P.: Using Convolutional Encoder-Decoder for Document Image Binarization. In: 14th International Conference on Document Analysis and Recognition (IAPR), (2017).
7. Kligler, N., Katz, S., Tal, A.: Document enhancement using visibility detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2374–2382 (2018).
8. Thomas, H., Qi, C., Deschaud, J., Marcotegui, B., Goulette, F., Guibas, L.: KPConv: Flexible and deformable convolution for point clouds. In: ICCV. (2019).
9. Zhuang, J.: Laddernet: Multi-path networks based on u-net for medical image segmentation. In: ArXiv (2018).
10. Pratikakis, L., Zagoris, K., Barlas, G., Gatos, B.: ICDAR 2017 Competition on Document Image Binarization DIBCO 2017. In: 14th IAPR International Conference on Document Analysis and Recognition, (2017).
11. Dang, Q.V, Lee, G.: Document Image Binarization Using Multi-scale Fusion Network. KIISE Journal, Vol. 46, pp. 1314-1321 (2019).

12. Gatos, B., Ntirogiannis, K., Pratikakis, I.: ICDAR 2009 document image binarization contest (DIBCO 2009). In: ICDAR, pp. 1375–1382 (2009).
13. Pratikakis, I., Gatos, B., Ntirogiannis, K.: H-DIBCO 2010 - handwritten document image binarization competition. In: ICFHR, pp. 727–732 (2010).
14. Pratikakis, I., Gatos, B., Ntirogiannis, K.: ICDAR 2011 document image binarization contest (DIBCO 2011). In: ICDAR, pp. 1506–1510 (2011).
15. Pratikakis, K., Gatos, B., Ntirogiannis, K.: ICFHR 2012 competition on handwritten document image binarization (H-DIBCO 2012). In: ICFHR, pp. 817–822 (2012).
16. Pratikakis, I., Gatos, B., Ntirogiannis, K.: ICDAR 2013 document image binarization contest (DIBCO 2013). In: ICDAR, pp. 1471–1476 (2013).
17. Pratikakis, I., Gatos, B., Ntirogiannis, K.: ICFHR 2014 competition on handwritten document image binarization (H-DIBCO 2014).In: ICFHR, pp. 809–813 (2014).
18. Pratikakis, I., Zagoris, K., Barlas, G., Gatos, B.: ICFHR2016 handwritten document image binarization contest (H-DIBCO 2016). In: ICFHR, pp. 619–623 (2016).
19. Otsu, N.: A threshold selection method from gray-level histograms. In: IEEE transactions on systems, man, and cybernetics 9 62–66 (1979).
20. Niblack, W.: An introduction to digital image processing. In: Strandberg Publishing Company (1985).