A Performance Improvement of Deep Learning Based Binarization of Degraded Document Images with the Use of the Voting Approach

Hubert Michalak[0000−0003−4888−4303] and Krzysztof Okarma[0000−0002−6721−3241]

West Pomeranian University of Technology in Szczecin
Faculty of Electrical Engineering
Department of Signal Processing and Multimedia Engineering
26 Kwietnia 10, 71-126 Szczecin, Poland

Abstract. In recent years a great progress and interest in the use of deep learning in computer vision applications may be observed, often leading to encouraging results. This also concerns image binarization methods, particularly for degraded document images where adaptive thresholding should be applied. Nevertheless, the use of neural networks, trained using relatively small datasets, may suffer from potential overfitting and the application of such trained deep networks for some other datasets does not always lead to satisfactory results. To increase the universality of such methods, the approach based on pixel voting has been proposed. In this approach, multiple methods, including those based on deep learning, are applied in parallel, and the final result depends on the majority of binarization results obtained using these methods at the pixel level. As verified for state-of-the-art datasets, the proposed approach leads to significant performance improvement in comparison to the other methods.

Keywords: Image binarization · Document images · Deep learning.

1 Introduction

Binarization of degraded document images, as well as natural images captured in uncontrolled lighting conditions, is still one of the active field of research in image processing and analysis. Such image preprocessing methods are useful not only for further text recognition in document images or the analysis of 2D binary codes but also for navigation of mobile robots, e.g., line followers, particularly in unstable lighting conditions. Therefore, highly degraded document images, particularly unevenly illuminated, are commonly accepted as benchmarks for newly proposed image binarization methods. Starting from well-known DIBCO datasets, recently some even more demanding datasets have been delivered, e.g. Bickley Diary, Nabuco and LiveMemory datasets or MonkCuper database. Some of them are available on the DIB website[1] hosted by Brazilian Universidade Federal de Pernambuco (UFPE).

[1] https://dib.cin.ufpe.br

Although for well-illuminated high quality images, Niblack-based adaptive thresholding methods, such as Sauvola, Wolf, Feng and NICK [1], or even global binarization methods, such as the most popular Otsu, might be enough, more demanding images require the development of more sophisticated solutions. One of the possible approaches is the use of deep neural networks, although such methods require time-consuming training and are not necessarily fast enough, as reported e.g. at the ICDAR 2021 Competition on Time-Quality Document Image Binarization [2]. Additionally, an important problem of such approaches may be their overfitting observed when the datasets used for training differ from testing images, considering image size, distortions, illumination, etc.

To demonstrate this issue, two deep-learning (DL) based methods have been used in the paper: the first developed by Sami Liedes[2] who trained the network using mainly DIBCO and Persian datasets, and the second, known as RObust document image BINarization tool (ROBIN), developed by Mikhail Masyagin[3]. Both tools have been written in Python in combination with some other open source projects such as OpenCV, Keras, Tensorflow or Augmentor, as reported in their documentation. An improvement of their performance, partially solving the overfitting problem, is possible using the pixel voting approach presented below.

2 Proposed Approach

Considering the progress in the development of modern processing units, there are wide possibilities of parallel processing of images, also using multiple methods independently. Hence, regardless of the relatively long computation time, observed for DL-based methods, their combination with some other approaches might be beneficial, potentially improving the obtained results, without significant increase of the total processing time. One of such approaches, considered in the paper, is the application of the pixel voting, successfully applied previously for some other image thresholding methods used for image preprocessing before the alphanumerical character recognition [5].

The main assumption of the pixel voting is the independent parallel binarization of an input image using a number of N selected algorithms where each pixel in the resulting image may be expressed as 0 or 1. Treating each ot these values as votes, the "winning" value may be selected as the final result for the considered pixel. Actually, this method may also be implemented simply as the median of the obtained binary values using each method at the pixel level. Nevertheless, as presented in the paper [5], satisfactory results may be obtained by the application of some algorithms based on various assumptions and the combination of similar algorithms, e.g. only Niblack-inspired adaptive methods, usually leads to worse performance, particularly for non-uniformly illuminated document images. Hence, the combination of the DL-based methods with some others seems to be an appropriate assumption for further investigation.

[2] https://github.com/sliedes/binarize

[3] https://github.com/masyagin1998/robin

3 Experimental Results

The verification of the proposed use of the pixel voting has been made utilizing commonly used datasets, including DIBCO 2009–2019, Bickley Diary, Persian, Nabuco (part 1 with 15 images and part 2 – 20 images), LiveMemory (20 images) and Monk Cuper Set (25 images). Since the verifications have been made calculating some typical metrics [6] such as F-Measure, Accuracy, Distance Reciprocal Distortion (DRD) and Misclassification Penalty Metric (MPM), only the images with known ground-truth (GT) binary images may be used for this purpose. To illustrate the obtained results in a relatively compact representative form, only the obtained F-Measure (FM) values are presented in Table 1.

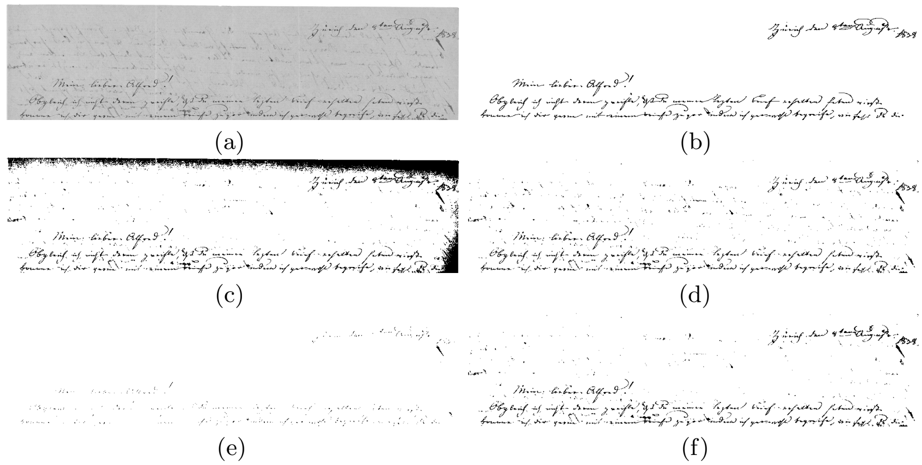
The previously proposed methods utilizing the stack of regions [3] are marked in Table 1 as 1L, 8L and 16L, according to the number of layers used in the calculations. The five sets of methods used in the pixel voting, the results for which are shown in Table 1, are the combinations of ROBIN with: 1L (single-layered) and JUCS (set 1), 16L and JUCS (set 2), Niblack and JUCS (set 3), 8L and Sauvola (set 4), and JUCS with Sauvola (set 5). The three best results for each dataset are marked by boldface fonts. The last row contains the results obtained for all datasets except DIBCO 2009–2018 (used for training of the DL-based methods). It may be easily noted that the best results obtained for DIBCO 2009–2018 datasets used for training are the effect of the overfitting.

Application of the deep-learning based methods for the other datasets, containing previously unseen data, leads to significantly worse results. Nevertheless, the use of the pixel voting, also in combination with some other previously proposed algorithms, makes it possible to improve the results of binarization.

Table 1. F-Measure values obtained for various datasets using some classical methods, the use of stack of regions, the JUCS method [4], deep learning methods and the proposed pixel voting.

Method	classical methods				stack of regions			JUCS	DL		Pixel Voting				
/ Dataset	Otsu	Nibl	Sauv	Brad	1L	8L	16L	[4]	ROBIN	Liedes	set 1	set 2	set 3	set 4	set 5
DIBCO 2009	78.6	76.8	78.7	77.0	77.1	81.3	81.4	84.8	**93.1**	71.8	87.1	87.0	86.7	**88.7**	**89.4**
DIBCO 2010	85.4	78.0	81.1	82.8	79.2	81.1	81.2	82.4	**94.2**	56.4	86.5	86.1	**87.1**	**86.7**	86.5
DIBCO 2011	82.1	68.9	78.7	74.9	72.6	75.4	75.5	81.0	**91.5**	65.0	83.8	83.4	83.3	**85.6**	**86.6**
DIBCO 2012	75.1	77.2	81.1	82.3	79.3	82.6	82.7	85.4	**94.7**	65.6	88.5	88.2	**89.0**	88.5	**88.9**
DIBCO 2013	80.0	72.7	78.8	77.5	76.1	78.9	78.9	82.7	**94.7**	71.4	86.3	85.9	86.3	**86.4**	**86.6**
DIBCO 2014	91.6	84.9	90.3	88.9	84.5	86.4	86.4	89.5	**96.0**	66.6	91.8	91.5	91.6	**93.7**	**94.5**
DIBCO 2016	86.6	74.2	80.1	76.0	76.3	81.1	81.2	86.1	**90.6**	68.3	87.5	87.5	87.2	**87.6**	**88.4**
DIBCO 2017	77.7	75.0	77.9	76.6	75.6	79.7	79.7	82.7	**92.2**	56.7	85.7	85.5	85.2	**86.3**	**86.9**
DIBCO 2018	51.5	67.6	54.6	61.0	68.0	70.3	70.4	72.5	**88.7**	57.6	**79.7**	78.5	**79.5**	76.4	75.7
DIBCO 2019A	**72.3**	54.4	48.1	54.3	58.4	60.8	61.0	71.1	46.9	26.6	**71.3**	**71.4**	70.8	65.4	70.5
DIBCO 2019B	23.3	54.2	44.4	42.1	52.1	54.4	54.4	58.2	43.3	34.2	**61.0**	**60.4**	**60.3**	59.0	59.6
Bickley Diary	58.8	83.8	72.4	70.6	78.9	84.4	84.5	83.6	73.1	42.8	**86.6**	**87.9**	**88.8**	86.5	83.2
Persian	82.1	78.2	**86.8**	81.2	79.2	83.0	83.1	85.8	85.5	72.6	86.6	86.7	86.0	**88.1**	**88.4**
Nabuco part 1	86.3	77.1	75.3	74.8	78.7	82.1	82.2	**87.1**	86.1	70.9	87.0	**87.2**	**87.4**	83.8	85.1
Nabuco part 2	**94.0**	82.8	83.3	81.0	84.6	88.6	88.6	**93.8**	90.2	71.9	92.9	**93.2**	93.1	90.0	91.3
Livememory	89.6	90.2	89.0	88.7	91.4	93.0	93.0	**94.8**	82.4	42.4	**93.0**	**93.9**	93.0	90.0	90.7
Monk Cuper Set	69.4	68.9	73.4	70.3	70.1	73.1	73.1	77.2	**84.8**	73.0	78.9	78.6	78.4	**79.0**	**79.8**
All	71.9	77.2	75.0	73.7	76.7	80.5	80.6	83.0	81.9	56.4	**85.2**	**85.5**	**85.8**	84.7	84.3
Verification	68.1	78.5	73.3	71.6	76.9	81.1	81.2	83.0	75.8	52.1	**84.7**	**85.3**	**85.6**	83.6	82.7

An illustration of the results obtained for a sample image from DIBCO2019 dataset, presenting the advantages of the pixel voting, is shown in Fig. 1.



(a) (b) (c) (d) (e) (f)

Fig. 1. Results obtained for a sample image from DIBCO2019 dataset track A, filename *02.bmp*: (a) input image, (b) ground truth, (c) Sauvola $FM = 36.12$, (d) JUCS $FM = 78.12$, (e) DL ROBIN $FM = 64.65$, (f) pixel voting – set 5 $FM = 81.72$.

References

1. Khurshid, K., Siddiqi, I., Faure, C., Vincent, N.: Comparison of Niblack inspired binarization methods for ancient documents. In: Document Recognition and Retrieval XVI. vol. 7247, pp. 7247–7247–9. SPIE (2009). https://doi.org/10.1117/12.805827
2. Lins, R.D., Bernardino, R.B., Smith, E.B., Kavallieratou, E.: ICDAR 2021 competition on time-quality document image binarization. In: Lladós, J., Lopresti, D., Uchida, S. (eds.) Document Analysis and Recognition – ICDAR 2021. pp. 708–722. Springer International Publishing, Cham (2021)
3. Michalak, H., Okarma, K.: Adaptive image binarization based on multi-layered stack of regions. In: Vento, M., Percannella, G. (eds.) Computer Analysis of Images and Patterns. LNCS, vol. 11679, pp. 281–293. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-29891-3_25
4. Michalak, H., Okarma, K.: Fast binarization of unevenly illuminated document images based on background estimation for optical character recognition purposes. Journal of Universal Computer Science **25**(6), 627–646 (2019). https://doi.org/10.3217/jucs-025-06-0627
5. Michalak, H., Okarma, K.: Robust combined binarization method of non-uniformly illuminated document images for alphanumerical character recognition. Sensors **20**(10), 2914 (may 2020). https://doi.org/10.3390/s20102914
6. Ntirogiannis, K., Gatos, B., Pratikakis, I.: Performance evaluation methodology for historical document image binarization. IEEE Transactions on Image Processing **22**(2), 595–609 (2013). https://doi.org/10.1109/TIP.2012.2219550