ICDAR 2021 Competition on Time-Quality Document Image Binarization

Rafael Dueire Lins[1,2](✉), Rodrigo Barros Bernardino[1], Elisa Barney Smith[3], and Ergina Kavallieratou[4]

[1] Universidade Federal de Pernambuco, Recife, PE 50730-120, Brazil
{rdl,rbb4}@cin.ufpe.br
[2] Universidade Federal Rural de Pernambuco, Recife, PE 52.171-900, Brazil
[3] Boise State University, Boise, ID, USA
ebarneysmith@boisestate.edu
[4] University of the Eagian, Mytilene, Greece
kavallieratou@aegean.gr

Abstract. The ICDAR 2021 Time-Quality Binarization Competition assessed the performance of 12 new and 49 other previously published binarization algorithms for scanned document images. Four test sets of "real-world" documents with different features were used. For each test set, the top twenty algorithms in the quality of the resulting two-tone images had their average processing time presented, yielding an account of their time complexity.

Keywords: Document binarization · DIB-dataset · Binarization competition

1 Introduction

Binarization algorithms take as input a color or grayscale image and produce a bi-tonal image as result. Binarization is a key preprocessing step in document analysis solutions such as automatic character recognition and layout analysis, as the quality of the monocromatic document has a strong effect on the performance for such document processing tasks. Besides that, document binarization can increase its readability and work as a file compression strategy [58], as the size of binary images is often orders of magnitudes smaller than the original gray or color images, requiring much less storage space or computer bandwidth for network transmission. Thus, assessing the quality of document binarization algorithms is of importance, a concern witnessed by contests that started more than a decade ago [14,44].

Any algorithm makes design assumptions about the nature of the data being processed. Document binarization algorithms are no exception. The DIB Team (https://dib.cin.ufpe.br/) presented evidence that "no binarization algorithm is good for all kinds of documents". Thus, depending on the intrinsic features of each document, a binarization algorithm may perform better than another.

J. Lladós et al. (Eds.): ICDAR 2021, LNCS 12824, pp. 708–722, 2021.
https://doi.org/10.1007/978-3-030-86337-1_47

The second point made by the DIB team is that "time matters". As binarization algorithms are part of several document processing pipelines, it is also important to balance the quality-time binomial. Several algorithms may produce binary images of similar quality with huge differences in time complexity.

Recently, a new series of document binarization contests were created not only comparing the enrolled participants among themselves, but also comparing the quality-time performance of the new with "classical" algorithms. Besides that, much larger datasets of "real-world" and synthetic documents were tested in different feature clusters. Scanned documents either historic (machine typed or handwritten, 200 or 300 dpi, with different hues and textures of paper, and with different degrees of back-to front interference) and bureaucratic (200 dpi, offset printed, and with light back-to-front interference) have their binarization quality-time assessed in contests [32,33].

This ICDAR 2021 competition assessed the quality-time performance of 12 new algorithms from seven teams from all over the world. The competitors' algorithms were also compared with the quality-time performance of 49 other binarization algorithms that may be considered "classic". Four test sets of "real-world" 200 and 300 dpi scanned documents were used in this assessment.

2 Participants

Seven research groups from six different countries, from the Americas, Asia, and Europe enrolled in this 2021 competition. In what follows, the teams are presented together with a brief outline of their approaches.

2.1 AutoHome Corp, Haidian Strict, Beijing, China Team: Huang Xiao, Liu Rong, Xu Chengshen, Li Lin, Ye Mingdeng

A combination of binary cross entropy and dice loss is chosen as the loss function of a deep-learning algorithm. Data augmentation is performed in the training process so as to improve the scores. The original colored or gray images are divided into patches with the same dimension (e.g. 128*128). For each colored patch, a trained Unet model is utilized to obtain a binarized patch. A binarized large image with the same size as the original image can be obtained with the combination of those binarized patches. In this method, the model stacking technique is performed via two Unet models with patch dimensions of 128*128 and 256*256. Further, a global view with a patch dimension of 512*512 is also combined to obtain the final results. The model with a global view is trained aiming to capture the global context and the character locations.

This team presented two different methods to this competition:

Method (1) - HuangBCD: The segmentation model is BCD-Unet based [2].

Method (2) - HuangUnet: The segmentation model is Unet based.

2.2 Universitas Syiah Kuala, Indonesia Team: Khairun Saddami

Method (1) - iNICK: An extension of the NICK binarization method [52]. The image standard deviation is used to determine the k value which is calculated as $k = -\sigma/(255 - 1.5\sigma)$, where σ is the image standard deviation that represents the image contrast.

Method (2) - CNW: Combination of Niblack and Wolf [49]. The threshold $T = (2m + mk((\sigma/m) - (\sigma/R) - 1))/2$, where σ is the image standard deviation, m is the mean of local window, R is the maximum standard deviation, $k = 0.35$.

Method (3) - CLD: Combined the local adaptive and global thresholding formulas, as described in [50].

2.3 West Pomeranian University of Technology, Poland Team: Hubert Michalak and Krzysztof Okarma

Method (1) - Michalak21_1: The first step of the algorithm is related to image downsampling where one of well-known interpolation methods may be applied. For this purpose the MATLAB function imresize was used with bilinear and the simple nearest neighbour method. Application of a relatively large kernel during downsizing of the image results in the loss of details related to shapes of individual characters. Therefore only the low frequency image data is preserved representing the overall distribution of the image brightness, being in fact mainly the downsampled background information. After resizing back the downsampled image to the original size using the same kernel, the image containing only the low frequency information is obtained, representing the approximated high resolution background. In the next step of the proposed method the subtraction of this image from the original is made to enhance the text data, followed by simple contrast increase and logical negation. The image obtained is subjected to the fast global thresholding using Otsu method.

Method (2) - Michalak21_2: The proposed method is based on the equalization of the illumination of an image, increasing also its contrast, making it easier to conduct the proper binarization. It is based on the analysis of the local entropy, assuming its noticeably higher values in the neighbourhood of the characters. Hence, only the relatively high entropy regions should be further analysed as potentially containing some characters, whereas low entropy regions may be considered as the background. The additional steps of the morphological dilation, increase of contrast and final binarization using Bradley's method are made during the final stage.

Method (3) - Michalak21_3: The initial idea of the application of the region based binarization for text recognition was presented assuming the application of document images containing predefined text. The proposed improved method assumes the division of the image into regions of NÎN pixels. For each of the regions the local threshold can be determined as $T = a * mean(X) - b$, where $mean(X)$ is the average brightness of the image region and the parameters a and b are subjected to optimization. The algorithm is based on the same idea of calculation of the local thresholds as the average brightness corrected by two parameters, however the number of regions is higher than would result from the resolution of the image and therefore they partially overlap each other. In this case for each sub-region several threshold values are calculated depending on the number of overlapping blocks covering the sub-region. The resulting local threshold is determined as the average of the threshold values calculated for the number of regions dependent on the assumed number of layers and the overlapping factor. The rationale for such an approach is a better tolerance of rapid illumination changes with the ability of a correct image binarization.

2.4 University of São Paulo (USP), Brazil Team: Diego Pavan Soler

The "DiegoPavan" binarization method chooses to downscale the input image, rather than using patching, and then rescaling the network output to the input original size. The network architecture used is based on DE-GAN [60], where we changed the input image to HSV, adjusted the hyperparameters and the training process, including image augmentation.

2.5 University of Fribourg, Sweden Team: Jean-Luc Bloechle

The "YinYang21" binarization algorithm detects the background of the original image using small overlapping windows. First, each window calculates its median color using a quantized color palette. Then, the estimated background image is generated by interpolating the computed median pixels of the overlapping windows. Next, the background image is subtracted from the original image, and the resulting difference image is transformed into grayscale, keeping only the lowest RGB component. A binarization is done by Otsu's algorithm. Detection and removal of small isolated connected components is made. The algorithm submitted in this competition is a faster and more accurate version of the previously submitted one in DocEng 2020 Binarization Competition [33].

2.6 Berlin State Library - Prussian Cultural Heritage Foundation, Germany Team: Vahid Rezanezhd, Clemens Neudecker and Konstantin Baierer

The "Vahid" algorithm is based on machine learning and it is in fact a pixel-wise segmentation model. The dataset used for training is a combination of training sets for binarization competitions in different years with pseudo-labeled images from our dataset in the Berlin State Library. A specific dataset has been produced for very dark or bright images. The model is based on a Resnet50-Unet [42].

2.7 Hubei University of Technology, China Team: Xinrui Wang, Wei Xiong, Min Li, Chuansheng Wang and Laifu Guan

The DocUNet method comprises three main steps. Firstly, a morphological bottom-hat transform is carried out to enhance the document image contrast, and the size of a disk-shaped structural element is determined by the stroke width transform (SWT). Secondly, a hybrid pyramid U-Net convolutional network [27] is performed on the enhanced document images for accurate pixel classification. Finally, the Otsu algorithm is applied as an image post-processing step to yield the final image.

3 Quality Evaluation Methods

To evaluate the binarization algorithms relative to image quality, the scanned documents were clustered according to their features (print type and paper texture luminosity). This produced five document sets. The quality of the binary images was compared using the PSNR, DRDM, F-Measure (FM) and pseudo-FMeasure (Fps) [41], and Cohen's Kappa [6,12]. The final ranking is defined by sorting the ranking summation in ascending order, following the methodology introduced by [46]. The consistency of the global ranking with a carefully made visual inspection was also checked.

Cohen's Kappa [12]

$$k = \frac{P_O - P_C}{1 - P_C}, \tag{1}$$

compares the observed accuracy with an expected accuracy, indicating how well a given classifier performs. P_O is the number of correctly mapped pixels (accuracy) and P_C is calculated by using:

$$P_C = \frac{n_{bf} \times n_{gf} + n_{bb} \times n_{gb}}{N^2}, \tag{2}$$

where n_{bf} and n_{bb} are the number of pixels mapped as foreground and background on the binary image, respectively, while n_{gf} and n_{gb} are the number of foreground and background pixels on the GT image and N is the total number of pixels. The Kappa coefficient has an excellent correspondence with the image-quality perception by human visual inspection of the resulting images. As indicated in [45], κ may be a good and easy to interpret image-quality evaluation measure for binary classifiers [6]. Thus, the top-twenty algorithms in image quality, ranked following [46], will have their κ coefficient and standard deviation (shown in parenthesis), together with the mean processing time and its standard deviation (also shown in parenthesis) presented in the tables of the results.

4 Processing-Time Evaluation

The 12 new algorithms assessed here were implemented by their authors. The purpose of the processing time evaluation here is to provide an order of magnitude of time elapsed for binarizing the whole datasets. The training-times for the AI-based algorithms were not considered. The competing 12 algorithms, printed in blue, were implemented using different programming languages, operating systems, and even for specific hardware platforms such as GPUs. They are compared against the other 49 algorithms in the literature, most of which were implemented by their authors or are available in image processing environments such as MatLab or ImageJ.

In the benchmaking, the following hardware was used:

- CPU: Intel(R) Core(TM) i7-10750H CPU @ 2.6 GHz RAM: 32 GB
- GPU: GeForce GTE 1650 4 GB

The algorithms were executed in the following platforms:

- **Windows 10 (version 1909):**
  - **Matlab:** Akbari_1 [1], Akbari_2 [1], Akbari_3 [1], ElisaTV [3], Ergina-Global [23], Ergina-Local [24], Gosh [5], Howe [18], Ghosh [5], Jia-Shi [20], Gattal [15], Lu-Su [36], Michalak [33], Yasin [32], iNICK [52], CNW [49], CLD [51], Michalak21_1, Michalak21_2, Michalak21_3.
- **Linux Pop!_OS 20.04**
  - **Python with GPU:** DilatedUNet [33]
  - **Python Machine Learning with CPU:** Calvo-Zaragoza [10], DiegoPavan [60], Doc-DLinkNet [68], DocUNet [32], HuangBCD, HuangUnet, Vahid, Yuleny [32].
  - **C++:** Bataineh [4], Bernsen [7], ISauvola [17], Niblack [40], Nick [25], Otsu [43], Sauvola [54], Singh [59], Su-Lu [61], WAN [39], Wolf [65].
  - **Java:** Bradley [9], dSLR [57], Huang [35], Intermodes [47], IsoData [63], Johannsen [21], KSW [22], Li-Tam [28], Mean [16], Mello-Lins [38], Min-Error [26], Minimum[47], Moments [62], Percentile [13], Pun [48], RenyEntropy [53], Shanbhag [56], Triangle [67], Wu-Lu [37], Yean-CC [66], YinYang [33], YinYang21.

Some experiments have been done with the algorithms that can be executed on both OS. No significant processing time difference was noticed. That is likely due to the fact that the exact same hardware and up to date compilers have been used in all cases. Although the user programming languages were Matlab, Python and Java, it is known that they are often used as an API to lower level implementations, leading to smaller differences in time purely due to the programming language used. Also, modern compilers are all nearly equally efficient. That can be easily verified by the results, as some algorithms using GPU performed fast, while others were slow. As for the Matlab implemented ones, some were among the fastest, while others among the slowest methods. Thus, if some optimization is made, using modern versions of the compilers, most algorithms should have similar performance in different languages.

The mean processing time was used in the analysis. It is most important to remark that the primary purpose here is to provide the order of magnitude of the processing time elapsed by each of the algorithms. The mean processing time figures are presented only to the 20 top **quality** algorithms, for each of the datasets tested.

5 Test Sets and Results

If one does not take into account some external physical noises such as stains, fungi, dirt that may affect document images [29], two aspects play fundamental role in complicating the binarization of scanned documents: the luminosity and texture of the paper background, that suffers the natural aging process, and the back-to-front interference. The back-to-front interference [31], later called bleeding or show-through, happens when a document is typed or written on both sides of a sheet of paper and the opacity of the paper is such as to allow the back printing or writing to be visualized on the front side. Given the importance of such a noise, most of the selected images from the Nabuco and LiveMemory datasets have such types of noise, with different degrees of interference α [30]. As already mentioned, four test sets of "real-world" 200 and 300 dpi scanned documents were used in this assessment.

The Nabuco and LiveMemory datasets used in the experiments here are part of the DIB - Document Image Binarization data set (https://dib.cin.ufpe.br/), which is part of the IAPR-TC10/TC11 open repository of document images [30].

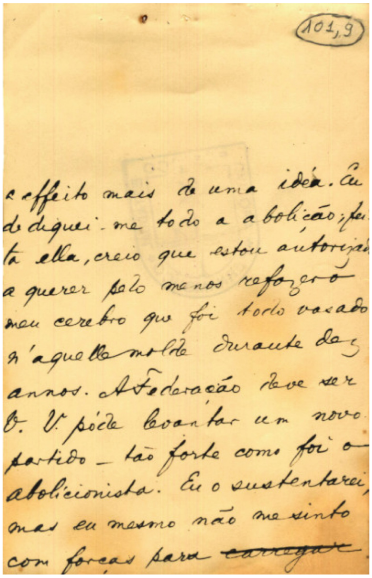
From the Nabuco bequest of historical documents from the late XIX century, 20 images have been selected, which were subdivided into three clusters according to the average luminosity level of the background texture. Dark textures have an average luminosity of 147, mid texture of 193 and light texture of 220. A total of seven dark, seven light texture handwritten and six mid-dark texture typewritten documents were selected. From the LiveMemory project, five images with various configurations were selected. From the PRImA project, four images that belong to the Europeana Newspapers Project Dataset were used. The images have been selected in order to provide some variability between the datasets, but similar images within the datasets. The chosen datasets are representative of a large number of "real-world" documents of interest.

The ground-truth images used here were obtained by binarizing the original images with the ten best quality algorithms from the previous competitions [32,33] in images similar to the ones chosen to this competition. Such images underwent a careful visual inspection. The three best binary images were merged by applying the AND logical operator. The resulting image underwent salt and pepper filtering. The resulting image was visually re-inspected and underwent an eventual manual cleaning.

5.1 Nabuco Dataset

The letters of Joaquim Nabuco (b. 1849/d. 1910), a Brazilian stateman who was the first Brazilian ambassador to the USA, and one of the most expressive figures in freeing black slaves in the Americas, are of great historical importance, and some of them are available in the DIB dataset. Three sets of images of the Nabuco bequest were used in the assessment here.

Table 1. Quality-time results for Nabuco, Light Texture, Handwritten Documents

#	Team	Kappa (SD)	Time (SD)	Example Image
1	Vahid	0.89 (0.06)	10.18 (4.49)	
2	HuangUnet	0.87 (0.13)	24.91 (7.91)	
3	Akbari_1 [1]	0.84 (0.21)	4.91 (1.98)	
4	HuangBCD	0.87 (0.10)	113.29 (35.16)	
5	Akbari_2 [1]	0.84 (0.21)	4.95 (2.12)	
6	Akbari_3 [1]	0.84 (0.21)	4.89 (1.99)	
7	Jia-Shi [20]	0.84 (0.21)	4.87 (1.99)	
8	Wolf [64]	0.86 (0.05)	0.06 (0.03)	
9	Sauvola [55]	0.86 (0.06)	0.04 (0.02)	
10	DocDLink [32]	0.81 (0.18)	55.60 (26.86)	
11	Yasin	0.83 (0.10)	1.18 (0.99)	
12	Gosh [5]	0.81 (0.15)	31.84 (16.58)	
13	Su-Lu [8]	0.85 (0.06)	0.41 (0.18)	
14	Lu-Su [36]	0.81 (0.12)	16.15 (7.06)	
15	Minimum [47]	0.84 (0.10)	0.01 (0.01)	
16	iNICK [52]	0.81 (0.11)	5.32 (4.09)	
17	DilatedUNet [33]	0.80 (0.12)	44.43 (15.47)	
18	Intermodes [47]	0.80 (0.11)	0.01 (0.00)	
19	Mello-Lins [38]	0.79 (0.21)	0.01 (0.00)	
20	ElisaTV [3]	0.76 (0.20)	2.41 (1.06)	

5.2 The LiveMemory Dataset

The LiveMemory Project [34] was a pioneering initiative to build a digital library of the entire collection of proceedings of the technical events of the Brazilian Telecommunications Society (SBrT) led by the first author of this paper, back in 2007. The real challenge was to scan all the printed-only volumes, semi-automatically index all the papers, enhance image quality, and to binarize the images in way such as to allow all the volumes to be stored in one single DVD, which was handed to all members of the SBrT. The documents were scanned in 200 dpi, true-color and stored using the jpeg file-format with standard (1% loss). The LiveMemory dataset is clearly the one with smaller variation among images, as they are all "modern" documents, offset printed and have an uniform background with some back-to-front interference.

Table 2. Quality-time results for Nabuco, Dark Texture, Handwritten Documents

#	Team	Kappa (SD)	Time (SD)	Example Image
1	Sauvola [55]	0.91 (0.04)	0.03 (0.00)	
2	Gosh [5]	0.89 (0.03)	20.97 (2.09)	
3	Wolf [64]	0.89 (0.03)	0.04 (0.00)	
4	DocDLink [32]	0.88 (0.05)	42.12 (2.31)	
5	HuangBCD	0.89 (0.02)	89.30 (7.21)	
6	Su-Lu [8]	0.90 (0.06)	0.32 (0.04)	
7	HuangUnet	0.89 (0.03)	19.81 (1.54)	
8	Yasin	0.89 (0.04)	0.82 (0.24)	
9	iNICK [52]	0.89 (0.03)	3.19 (0.51)	
10	Nick [25]	0.89 (0.03)	0.03 (0.00)	
11	Singh [59]	0.89 (0.03)	0.03 (0.00)	
12	YinYang21	0.86 (0.07)	0.51 (0.07)	
13	DocUNet [32]	0.85 (0.07)	37.33 (4.53)	
14	Li-Tam [28]	0.86 (0.05)	0.01 (0.00)	
15	Vahid	0.86 (0.06)	7.39 (0.49)	
16	Shanbhag [56]	0.85 (0.10)	0.01 (0.00)	
17	Howe [18]	0.85 (0.07)	15.59 (7.72)	
18	DilatedUNet [33]	0.85 (0.07)	31.96 (3.14)	
19	Ergina_L [24]	0.86 (0.07)	0.12 (0.02)	
20	Ergina_G [23]	0.85 (0.08)	0.08 (0.01)	

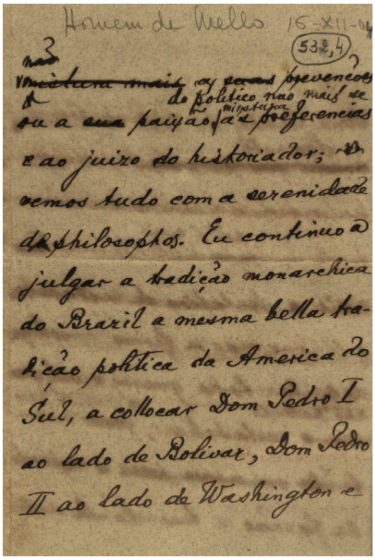
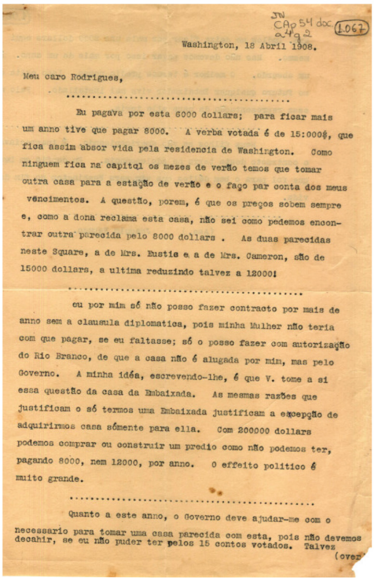

Homem de Mello 16-XII-04
532,4
3
~~não vinham mais~~ as ~~suas~~ prevenções
de politico não mais se
ou a ~~sua~~ paixão, já as preferencias
e ao juizo do historiador;
vemos tudo com a serenidade
dos philosophos. Eu continuo a
julgar a tradição monarchica
do Brazil a mesma bella tra-
dição politica da America do
Sul, a collocar Dom Pedro I
ao lado de Bolivar, Dom Pedro
II ao lado de Washington e

Table 3. Quality-time results for Nabuco, Mid Texture, Typewritten Documents

#	Team	Kappa (SD)	Time (SD)	Example Image
1	Gosh [5]	0.92 (0.07)	51.82 (6.28)	
2	HuangUnet	0.91 (0.05)	37.67 (1.81)	
3	Yasin	0.90 (0.06)	1.03 (0.14)	
4	HuangBCD	0.91 (0.04)	167.59 (7.49)	
5	iNICK [52]	0.89 (0.07)	3.70 (0.52)	
6	Wolf [64]	0.92 (0.03)	0.10 (0.01)	
7	Singh [59]	0.92 (0.04)	0.13 (0.01)	
8	Michalak21a	0.87 (0.10)	0.02 (0.00)	
9	Li-Tam [28]	0.88 (0.08)	0.02 (0.00)	
10	Minimum [47]	0.90 (0.03)	0.02 (0.00)	
11	Nick [25]	0.91 (0.05)	0.08 (0.00)	
12	Su-Lu [8]	0.91 (0.02)	0.71 (0.07)	
13	Intermodes [47]	0.87 (0.06)	0.02 (0.00)	
14	Michalak21c	0.85 (0.10)	0.47 (0.04)	
15	ElisaTV [3]	0.86 (0.08)	4.27 (0.20)	
16	Akbari_1 [1]	0.86 (0.06)	8.45 (0.85)	
17	Akbari_2 [1]	0.86 (0.06)	8.45 (0.87)	
18	Bradley [9]	0.84 (0.09)	0.14 (0.01)	
19	Akbari_3 [1]	0.86 (0.06)	8.46 (0.87)	
20	Jia-Shi [20]	0.86 (0.06)	8.46 (0.88)	


JW
CAp 54 doc.
a4q2
1.067

Washington, 18 Abril 1908.

Meu caro Rodrigues,

....

Eu pagava por esta 6000 dollars; para ficar mais um anno tive que pagar 8000. A verba votada é de 15:000$, que fica assim absor vida pela residencia de Washington. Como ninguem fica na capitql os mezes de verão temos que tomar outra casa para a estação de verão e o faço par conta dos meus vencimentos. A questão, porem, é que os preços sobem sempre e, como a dona reclama esta casa, não sei como pedemos encontrar outra parecida pelo 8000 dollars . As duas parecidas neste Square, a de Mrs. Eustis e a de Mrs. Cameron, são de 15000 dollars, a ultima reduzindo talvez a 12000!

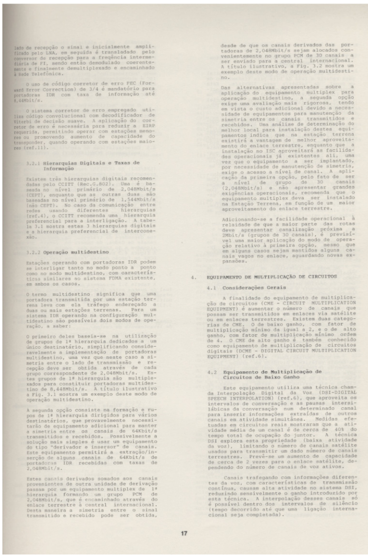
....

eu por mim só não posso fazer contracto por mais de anno sem a clausula diplomatica, pois minha Mulher não teria com que pagar, se eu faltasse; só o posso fazer com autorização do Rio Branco, de que a casa não é alugada por mim, mas pelo Governo. A minha idéa, escrevendo-lhe, é que V. tome a si essa questão da casa da Embaixada. As mesmas razões que justificam o só termos uma Embaixada justificam a excepção de adquirirmos casa sómente para ella. Com 200000 dollars podemos comprar ou construir um predio como não podemos ter, pagando 8000, nem 12000, por anno. O effeito politico é muito grande.

....

Quanto a este anno, o Governo deve ajudar-me com o necessario para tomar uma casa parecida com esta, pois não devemos decahir, se eu não puder ter pelos 15 contos votados. Talvez
(over

Table 4. Quality-time results for LiveMemory Test Set

#	Team	Kappa (SD)	Time (SD)	Example Image
1	Michalak [33]	0.94 (0.04)	0.08 (0.05)	
2	Bradley [9]	0.94 (0.05)	0.29 (0.01)	
3	Wolf [64]	0.94 (0.05)	0.22 (0.02)	
4	ElisaTV [3]	0.93 (0.06)	9.55 (1.15)	
5	Gosh [5]	0.94 (0.03)	111.80 (21.29)	
6	IsoData [63]	0.90 (0.12)	0.14 (0.02)	
7	Gattal [15]	0.91 (0.11)	54.40 (1.48)	
8	Otsu [43]	0.90 (0.13)	0.02 (0.00)	
9	Li-Tam [28]	0.91 (0.10)	0.14 (0.01)	
10	Yasin	0.93 (0.05)	2.05 (0.99)	
11	iNICK [52]	0.93 (0.03)	3.48 (0.35)	
12	Michalak21_1	0.94 (0.03)	0.08 (0.05)	
13	Intermodes [47]	0.92 (0.07)	0.14 (0.02)	
14	Michalak21_3	0.92 (0.06)	1.32 (0.67)	
15	Johannsen [21]	0.92 (0.05)	0.14 (0.02)	
16	Su-Lu [8]	0.93 (0.02)	1.67 (0.10)	
17	YinYang21	0.91 (0.07)	1.60 (0.13)	
18	HuangBCD	0.92 (0.07)	316.87 (25.66)	
19	HuangUnet	0.92 (0.07)	316.78 (26.17)	
20	WAN [39]	0.92 (0.07)	1.01 (0.09)	

5.3 PRImA

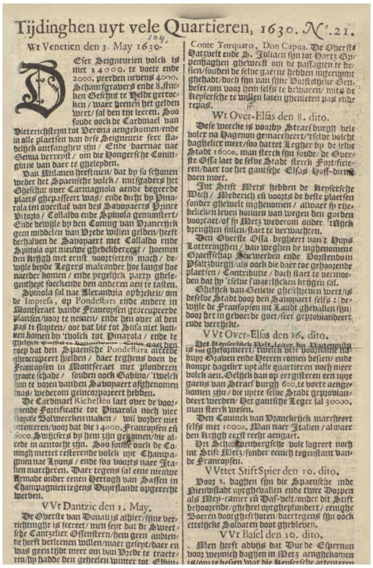
Europeana Newspapers: [11] Its main goal is to provide a representative collection of all the types of newspapers which are and/or might be subject of ongoing or future digitisation activities. As such, it is hosting scanned images, metadata and ground truth (a representation of the ideal result of a processing step like OCR or layout analysis) on the level of individual newspaper pages. Three of the selected images are from the Royal Library, in the Netherlands, and one is from the Austrian National Library, Austria.

6 Result Analysis

This contest is designed to look at the tradeoff between binarization performance and computational time. There is no single winner. The 20 top performing teams are reported by dataset in Tables 1, 2, 3, 4 and 5. Yasin and HuangBCD appeared in the top ranked algorithms for all five datasets and the sister algorithm HuangUnet appeared in the top ranked for four of the datasets. Michalak21's first and third algorithms both appeared three times in the rankings. YinYang21 appeared 3 times and Vahid appeared twice.

The average kappa values for the top 20 reported for each dataset fell in a narrow range from 0.75 to 0.94, The binarized images produced using the best quality algorithm for the test images, as one may expect, had very high

Table 5. Quality-time results for PRImA Data Set

#	Team	Kappa (SD)	Time (SD)	Example Image (cropped)
1	Gosh [5]	0.90 (0.09)	159.77 (92.16)	
2	Bradley [9]	0.90 (0.08)	0.43 (0.35)	
3	Michalak21a	0.89 (0.08)	0.10 (0.06)	
4	Intermodes [47]	0.89 (0.14)	0.19 (0.23)	
5	Michalak [33]	0.91 (0.08)	0.10 (0.06)	
6	Li-Tam [28]	0.87 (0.17)	0.19 (0.23)	
7	DocDLink [32]	0.92 (0.06)	292.46 (223.60)	
8	ElisaTV [3]	0.88 (0.04)	13.56 (10.63)	
9	IsoData [63]	0.87 (0.14)	0.19 (0.23)	
10	Su-Lu [8]	0.87 (0.10)	2.93 (1.95)	
11	Moments [62]	0.85 (0.16)	0.19 (0.23)	
12	Michalak21c	0.89 (0.05)	1.83 (1.05)	
13	Yasin	0.87 (0.08)	2.42 (1.24)	
14	Ergina_L [24]	0.87 (0.10)	1.28 (0.68)	
15	Gattal [15]	0.87 (0.13)	56.94 (3.80)	
16	Akbari_1 [1]	0.86 (0.06)	32.40 (20.71)	
17	Ergina_G [23]	0.86 (0.14)	0.85 (0.62)	
18	Huang [19]	0.86 (0.10)	0.19 (0.22)	
19	Akbari_2 [1]	0.86 (0.06)	32.39 (20.68)	
20	HuangBCD	0.86 (0.08)	445.08 (301.37)	

Tijdinghen uyt vele Quartieren, 1630. N°. 21.

visual quality. The ten top quality images for each of the sample images will be made available at the DIB website (https://dib.cin.ufpe.br/) immediately after ICDAR 2021.

The execution times varied more significantly than the performance as measured by kappa value. The median run time of the top performing algorithms was 1 s with 21% of the algorithms taking less than 0.1 ss. Michalak21_1 was the fastest of the ranked new algorithms competing this year. Nine of the algorithms took more than a minute on average to process the page, which for most applications will not be practical for the small performance benefit the algorithm may offer. HuangBCD that appeared in the top rankings for all datasets was also the algorithm that had the longest run time of all the ranked algorithms. The median run time for the algorithms published before 2010 was 0.11 s, whereas the median of the algorithms published 2010–2019 increased to 4.95 s and the median of those published in 2020 and 2021 is 7.39. The performance does not vary significantly between those groups.

7 Conclusions

This ICDAR 2021 Competition on Time-Quality Document Image Binarization shows that document image binarization is still a challenging task. The number of ways the problem can be made more difficult leads to demand to develop a new algorithm that can handle that one outlier case which others could not properly binarize. Machine-learning binarization algorithms are rising in providing better quality images, but some of the classic algorithms like IsoData [63] and Savoula [55] continued to appear in the top ranked algorithm list and they still provide very good, if not the best quality bitonal-image at a much lower time complexity.

Machine-learning binarization algorithms are rising in providing better quality images, but some of the classic algorithms still provide very good, if not the best quality bitonal-image at a much lower time complexity. It is important to remark that the training-time for the machine-learning based algorithms was not computed. Another point worth remarking is that some of those ML algorithms require computational resources that may be considered prohibitive, as some of the competing algorithms in the ICDAR 2019 Competition on Time-Quality Document Image Binarization [32] were still unable to run to all test images of the test sets used here.

Acknowledgements. The authors grateful to all the participants of this ICDAR competition, and to all researchers who made the code for their binarization algorithms publically available. This research was partly sponsored by CNPq − Brazilian Government.

References

1. Akbari, Y., Britto Jr., A.S., Al-Maadeed, S., Oliveira, L.S.: Binarization of degraded document images using convolutional neural networks based on predicted two-channel images. In: ICDAR (2019)
2. Azad, R., Asadi-Aghbolaghi, M., Fathy, M., Escalera, S.: Bi-directional ConvLSTM U-net with densley connected convolutions. In: ICCVW 2019, pp. 406–415 (2019)
3. Barney Smith, E.H., Likforman-Sulem, L., Darbon, J.: Effect of pre-processing on binarization. In: Document Recognition and Retrieval XVII, p. 75340H (2010)
4. Bataineh, B., Abdullah, S.N.H.S., Omar, K.: An adaptive local binarization method for document images based on a novel thresholding method and dynamic windows. Pattern Recogn. Lett. **32**(14), 1805–1813 (2011)
5. Bera, S.K., Ghosh, S., Bhowmik, S., Sarkar, R., Nasipuri, M.: A non-parametric binarization method based on ensemble of clustering algorithms. Multimed. Tools Appl. **80**(5), 7653–7673 (2020). https://doi.org/10.1007/s11042-020-09836-z
6. Bernardino, R., Lins, R., Jesus, D.M.: A quality and time assessment of binarization algorithms. In: 2019 15th IAPR ICDAR, pp. 1444–1450 (2019)
7. Bernsen, J.: Dynamic thresholding of gray-level images. In: International Conference on Pattern Recognition, pp. 1251–1255 (1986)
8. Su, B., Lu, S., Tan, C.L.: Robust document image binarization technique for degraded document images. Trans. Image Process. **22**(4), 1408–1417 (2013)
9. Bradley, D., Roth, G.: Adaptive thresholding using the integral image. J. Graph. Tools **12**(2), 13–21 (2007)
10. Calvo-Zaragoza, J., Gallego, A.J.: A selectional auto-encoder approach for document image binarization. Pattern Recogn. **86**, 37–47 (2019)
11. Clausner, C., Papadopoulos, C., Pletschacher, S., Antonacopoulos, A.: The ENP image and ground truth dataset of h. newspapers. In: ICDAR'15, pp. 931–935 (2015)

12. Congalton, R.G.: A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens. Environ. **37**(1), 35–46 (1991)
13. Doyle, W.: Operations useful for similarity-invariant pattern recognition. J. ACM **9**(2), 259–267 (1962)
14. Gatos, B., Ntirogiannis, K., Pratikakis, I.: ICDAR 2009 document image binarization contest (DIBCO 2009). In: 2009 10th ICDAR, pp. 1375–1382. IEEE (2009)
15. Gattal, A., Abbas, F., Laouar, M.R.: Automatic parameter tuning of K-means algorithm for document binarization. In: 7th ICSENT, pp. 1–4. ACM Press (2018)
16. Glasbey, C.: An analysis of histogram-based thresholding algorithms. Graph. Model. Image Process. **55**(6), 532–537 (1993)
17. Hadjadj, Z., Meziane, A., Cherfa, Y., Cheriet, M., Setitra, I.: ISauvola: improved Sauvola's algorithm for document image binarization. In: Campilho, A., Karray, F. (eds.) ICIAR 2016. LNCS, vol. 9730, pp. 737–745. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-41501-7_82
18. Howe, N.R.: Document binarization with automatic parameter tuning. IJDAR **16**(3), 247–258 (2013). https://doi.org/10.1007/s10032-017-0293-7
19. Huang, L.K., Wang, M.J.J.: Image thresholding by minimizing the measures of fuzziness. Pattern Recogn. **28**(1), 41–51 (1995)
20. Jia, F., Shi, C., He, K., Wang, C., Xiao, B.: Degraded document image binarization using structural symmetry of strokes. Pattern Recogn. **74**, 225–240 (2018)
21. Johannsen, G., Bille, J.: A threshold selection method using information measures. In: International Conference Pattern Recognition, pp. 140–143 (1982)
22. Kapur, J., Sahoo, P., Wong, A.: A new method for gray-level picture thresholding using the entropy of the histogram. Comput. Vis. Graph. Image Process. **29**(1), 140
23. Kavallieratou, E.: A binarization algorithm specialized on document images and photos. In: ICDAR 2005, no. 1, pp. 463–467 (2005)
24. Kavallieratou, E., Stathis, S.: Adaptive binarization of historical document images. In: Proceedings - International Conference on Pattern Recognition, vol. 3, pp. 742–745 (2006)
25. Khurshid, K., Siddiqi, I., Faure, C., Vincent, N.: Comparison of Niblack inspired binarization methods for ancient documents. In: SPIE, p. 72470U (2009)
26. Kittler, J., Illingworth, J.: Minimum error thresholding. Pattern Recogn. **19**(1), 41–47 (1986)
27. Kong, X., Sun, G., Wu, Q., Liu, J., Lin, F.: Hybrid pyramid U-Net model for brain tumor segmentation. In: Shi, Z., Mercier-Laurent, E., Li, J. (eds.) IIP 2018. IAICT, vol. 538, pp. 346–355. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00828-4_35
28. Li, C., Tam, P.: An iterative algorithm for minimum cross entropy thresholding. Pattern Recogn. Lett. **19**(8), 771–776 (1998)
29. Lins, R.D.: A taxonomy for noise in images of paper documents - the physical noises. In: Kamel, M., Campilho, A. (eds.) ICIAR 2009. LNCS, vol. 5627, pp. 844–854. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02611-9_83
30. Lins, R.D., Almeida, M.M.D., Bernardino, R.B., Jesus, D., Oliveira, J.M.: Assessing binarization techniques for document images. In: DocEng'17, pp. 183–192 (2017)
31. Lins, R.D., Guimarães Neto, M., França Neto, L., Galdino Rosa, L.: An environment for processing images of historical documents. Microprocess. Microprogram. **40**(10–12), 939–942 (1994)

32. Lins, R.D., Kavallieratou, E., Barney Smith, E., Bernardino, R.B., de Jesus, D.M.: ICDAR 2019 Time-Quality Binarization Competition. In: ICDAR, pp. 1539–1546 (2019)
33. Lins, R.D., Simske, S.J., Bernardino, R.B.: DocEng'2020 time-quality competition on binarizing photographed documents. In: DocEng'20, pp. 1–4. ACM (2020)
34. Lins, R.D., Torreão, G., Pereira e Silva, G.: Content recognition and indexing in the LiveMemory platform. In: Ogier, J.-M., Liu, W., Lladós, J. (eds.) GREC 2009. LNCS, vol. 6020, pp. 220–230. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-13728-0_20
35. Lu, D., Huang, X., Sui, L.X.: Binarization of degraded document images based on contrast enhancement. IJDAR **21**(1–2), 123–135 (2018)
36. Lu, S., Su, B., Tan, C.L.: Document image binarization using background estimation and stroke edges. IJDAR **13**(4), 303–314 (2010)
37. Lu, W., Songde, M., Lu, H.: An effective entropic thresholding for ultrasonic images. In: 14th ICPR, vol. 2, pp. 1552–1554 (1998)
38. Mello, C.A.B., Lins, R.D.: Image segmentation of historical documents. In: Visual 2000 (2000)
39. Mustafa, W.A., Abdul Kader, M.M.M.: Binarization of document image using optimum threshold modification. J. Phys.: C. Ser. **1019**(1), 012022 (2018)
40. Niblack, W.: An Introduction to Digital Image Processing. Strandberg Publishing Company, Copenhagen (1985)
41. Ntirogiannis, K., Gatos, B., Pratikakis, I.: Performance evaluation methodology for historical document image binarization. IEEE Trans. Image Process. **22**(2), 595–609 (2013)
42. Oliveira, S.A., Seguin, B., Kaplan, F.: dhSegment: a generic deep-learning approach for document segmentation. CoRR abs/1804.1 (2018)
43. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. **9**(1), 62–66 (1979)
44. Paredes, R., Kavallieratou, E., Lins, R.D.: ICFHR 2010 contest: quantitative evaluation of binarization algorithms. In: 12th ICFHR, pp. 733–736. IEEE (2010)
45. Powers, D.M.W.: Evaluation: from precision, recall and F-measure to ROC, informedness, markedness & correlation. J. M. L. Technol. **2**(1), 37–63 (2011)
46. Pratikakis, I., et al.: ICDAR 2019 competition on document image binarization. In: ICDAR, No. November 2019 (2019)
47. Prewitt, J.M.S., Mendelsohn, M.L.: The analysis of cell images. Ann. N. Y. Acad. Sci. **128**(3), 1035–1053 (2006)
48. Pun, T.: Entropic thresholding, a new approach. Comput. Graph. Image Process. **16**(3), 210–239 (1981)
49. Saddami, K., Afrah, P., Mutiawani, V., Arnia, F.: A new adaptive thresholding technique for binarizing ancient document. In: INAPR, pp. 57–61. IEEE (2018)
50. Saddami, K., Munadi, K., Away, Y., Arnia, F.: Combination local and global thresholding method for binarizing ancient Jawi document. JTIIK (2019)
51. Saddami, K., Munadi, K., Away, Y., Arnia, F.: Effective and fast binarization method for combined degradation on ancient documents. Heliyon (2019)
52. Saddami, K., Munadi, K., Muchallil, S., Arnia, F.: Improved thresholding method for enhancing Jawi binarization performance. In: ICDAR, vol. 1, pp. 1108–1113. IEEE (2017)
53. Sahoo, P., Wilkins, C., Yeager, J.: Threshold selection using Renyi's entropy. Pattern Recogn. **30**(1), 71–84 (1997)
54. Sauvola, J., Pietikäinen, M., Pietikainem, M.: Adaptive document image binarization. Pattern Recogn. **33**(2), 225–236 (2000)

55. Sauvola, J., Seppanen, T., Haapakoski, S., Pietikainen, M.: Adaptive document binarization. In: ICDAR, vol. 1, pp. 147–152. IEEE Comput. Soc (1997)
56. Shanbhag, A.G.: Utilization of information measure as a means of image thresholding. CVGIP: Graph. Models Image Process. **56**(5), 414–419 (1994)
57. Silva, J.M.M., Lins, R.D., Rocha, V.C.: Binarizing and filtering historical documents with back-to-front interference. ACM SAC **2006**, 853–858 (2006)
58. da Silva, J.M.M., Lins, R.D.: Color document synthesis as a compression strategy. In: ICDAR (ICDAR), pp. 466–470 (2007)
59. Singh, T.R., Roy, S., Singh, O.I., Sinam, T., Singh, K.M.: A new local adaptive thresholding technique in binarization. IJCSI Int. J. Comput. Sci. Issues **08**(6), 271–277 (2011)
60. Souibgui, M.A., Kessentini, Y.: DE-GAN: a conditional generative adversarial network for document enhancement. Trans. Pattern Anal. Mach. Intell. 1 (2021)
61. Su, B., Lu, S., Tan, C.L.: Binarization of historical document images using the local maximum and minimum. In: 8th IAPR DAS, pp. 159–166. ACM Press (2010)
62. Tsai, W.H.: Moment-preserving thresolding: a new approach. Comput. Vis. Graph. Image Process. **29**(3), 377–393 (1985)
63. Velasco, F.R.: Thresholding using the Isodata clustering algorithm. Technical report, OSD or Non-Service DoD Agency (1979)
64. Wolf, C., Doermann, D.: Binarization of low quality text using a Markov random field model. In: Object Recognition Supported by User Interaction for Service Robots, vol. 3, pp. 160–163. IEEE Computer Society (2002)
65. Wolf, C., Jolion, J.M., Chassaing, F.: Text localization, enhancement and binarization in multimedia documents. In: Object Recognition Supported by User Interaction for Service Robots, vol. 2, pp. 1037–1040. IEEE Computer Society (2003)
66. Yen, J.C., Chang, F.J.C.S., Yen, J.C., Chang, F.J., Chang, S.: A new criterion for automatic multilevel thresholding. Trans. Image Process. **4**(3), 370–378 (1995)
67. Zack, G.W., Rogers, W.E., Latt, S.A.: Automatic measurement of sister chromatid exchange frequency. J. Histochem. Cytochem. **25**(7), 741–753 (1977)
68. Zhou, L., Zhang, C., Wu, M.: D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. In: Computer Vision and Pattern Recognition (2018)