ICDAR 2019 Competition on Document Image Binarization (DIBCO 2019)

Ioannis Pratikakis[1], Konstantinos Zagoris[1], Xenofon Karagiannis[1], Lazaros Tsochatzidis[1], Tanmoy Mondal[2] and Isabelle Marthot-Santaniello[3]

[1]Visual Computing Group, Department of Electrical and Computer Engineering, Democritus University of Thrace, 67100 Xanthi, Greece, e-mail: {ipratika, kzagoris, xkaragia, ltsochat} @ee.duth.gr

[2]Zenith Team, INRIA, Montpellier, France
e-mail: tanmoy.mondal@inria.fr

[3]PI d-scribes project, Universität Basel, Switzerland
e-mail: i.marthot-santaniello@unibas.ch

***Abstract* — DIBCO 2019 is the international Competition on Document Image Binarization organized in conjunction with the ICDAR 2019 conference. The general objective of the contest is to identify current advances in document image binarization of machine-printed and handwritten document images using performance evaluation measures that are motivated by document image analysis and recognition requirements. This paper describes the competition details including the evaluation measures used as well as the performance of the 24 submitted methods along with a brief description of each method.**

Keywords – machine-printed, handwritten document image, binarization, performance evaluation

I. INTRODUCTION

Document image binarization is of great importance in the document image analysis and recognition pipeline since it affects further stages of layout analysis and the recognition process. The evaluation of a binarization method aids in verifying its effectiveness and studying its algorithmic behaviour. In this respect, it is imperative to create a framework for benchmarking purposes, i.e. a benchmarking dataset along with an objective evaluation methodology in order to capture the efficiency of current image binarization practices for document images. In view of this, the DIBCO series competition is active since 2009 [1] which is dedicated to benchmarking binarization algorithms of both machine-printed and handwritten document images. The most recent examples are the H-DIBCO 2018 [2] organized in conjunction with ICFHR 2018 and the DIBCO 2017 [3] organized in conjunction with ICDAR 2017.

In this paper, we present the results of DIBCO 2019, organized in conjunction with ICDAR 2019 using document images with various complexity for which we created the binary image ground truth. The authors of submitted methods registered in the competition and downloaded representative document images along with the corresponding ground truth from previous DIBCO contests available in the competition's site (https://vc.ee.duth.gr/dibco2019/). In the sequel, all registered participants were required to submit their method's executable program. After the evaluation of all candidate methods, the testing dataset which comprises a total of 20 images described in Section IV, the associated ground truth as well as the evaluation software are publicly available at: http://vc.ee.duth.gr/dibco2019/benchmark.

II. METHODS AND PARTICIPANTS

Fifteen (15) research teams have participated in the competition with twenty four (24) distinct algorithms (Participant 5 submitted three algorithms while Participants 1, 2, 3, 7, 8, 10, 14 submitted two algorithms). Brief descriptions of the methods are given in the following (the order of appearance is the chronological order of the algorithm's submission).

1) Hubei University of Technology, Wuhan, P.R. China

Method a – authored by (*Xinrui Wang, Wei Xiong, Min Li, Chuansheng Wang, Laifu Guan*)

The submitted method, referred to as Doc-UNet, comprises three main steps. Firstly, 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) [4]. Secondly, a hybrid pyramid U-Net convolutional network [5] is performed on the enhanced document images for accurate pixel classification. Finally, the OTSU algorithm [6] is adopted as an image post-processing step.

Method b – authored by (*Xiuhong Jia, Wei Xiong, Jingyi Jin, Zijie Xiong, Min Li*)

The method, called Doc-DLinkNet, consists of three main steps. First, considering the restriction and various sizes of document images, the original image is cropped into 256×256 patches, and some data augmentation strategies such as shape shift and color shift are applied. Second, a D-LinkNet architecture [7] is adopted and trained by using document image patches as input and the corresponding binary maps as ground truths. D-LinkNet is a semantic segmentation neural network, which involves dilated convolution and pre-trained encoder. Finally, the Principal Components Analysis (PCA) method is used to perform image dimensionality reduction and feature extraction, and then generate the final results according to optimal parameters learned from the training procedure.

2) Chonnam National University, Gwangju, Korea (*Quang-Vinh Dang, Xuan-Bac Nguyen, Gueesang Lee*)

Two binarization methods have been submitted that are based on LadderNet [8]. In particular,

Method a:

A properly designed LadderNet architecture Model is considered that is trained independently using document image patches (48x48) as input and binary maps as ground truth. In testing stage, a larger size of striding window is used

Method b:

This relies upon the combination of two properly designed LadderNet architectures wherein one is built as a deep architecture and the other is built as a comparatively shallower architecture. Each structure is trained independently with image patches (48x48). In testing stage, it has been observed that in the case of deeper architectures with larger size of striding window while the predicted maps produce less noise in the background the detail of the text is not clear. In the case of shallower architectures with smaller size of striding window the predicted maps have improved text strokes but contain more background noise. In view of this, a better result is achieved by combining the outputs of these two architectures.

3) Autohome Inc, Beijing, China, P.R. China (*Huang Xiao, Liu Rong, Xu Chengshen, Li Lin*)

The document image binarization task is considered as an image segmentation problem via a deep-learning algorithm based on the Unet popular network. A combination of binary cross entropy and dice loss is chosen as the loss function. Data augmentation is performed in the training process so as to improve the scores. In this competition, two methods are provided as follows:

Method a:

The original colored or gray images are divided into patches with the same dimension (e.g. 128*128). For each patch, a trained Unet model is utilized to obtain a binarized patch. Finally, 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, a model stacking technique is performed via two Unet models with patch dimension 128*128 and 256*256, respectively.

Method b:

It is an extension that is based on Method (a). A global view with patch dimension 512*512 and a local view provided in Method (a) is combined. The Unet model with a global view is trained aiming at capturing the global context and the character locations. The Unet model in Method (a) is trained in a small scale which is easier to precisely outline the boundaries of the characters.

4) University of Groningen, the Netherlands (*Sheng He, Lambert Schomaker*)

The method relies upon training the neural network to learn the degradations in document images and produce uniform images of the degraded input images, which in turn allows the network to refine the output iteratively. The stacked refinement (SR) is applied, which uses a stack of different neural networks for iterative output refinement. Given the learned nature of the uniform and enhanced image, the binarization map can be easily obtained through the use of a global Otsu threshold. A detailed description of the method is given in [9].

5) Qatar University, Qatar and Pontifícia Universidade Católica do Paraná (PUCPR), Curitiba, PR, Brazil and Universidade Federal do Paraná (UFPR), Curitiba, PR, Brazil (*Younes Akbari, Alceu S. Britto Jr., Somaya Al-maadeed, Luiz S. Oliveira*)

The core document image binarisation methodology is relied upon a Segnet network architecture which is fed by multichannel images that correspond to the original, the image approximations based on the coefficients of the three wavelet subbands and the binarized image produced by the structural symmetric pixels (SSPs) method [10]. Taking into account the aforementioned images two types of multichannel images have been used. In the first type, the multichannel image comprises four-channels and has been constructed from the original image and an image approximation based on the coefficients of three sub-bands. In the second type, a two-channel image is considered that originates from the original image and its binarized counterpart using the structural symmetric pixels (SSPs) method [11]. At the level of network architecture, two approaches have been considered: single and multiple networks. In the case of a single network, the input image consists of the original image (grayscale mode) together with all subbands (three subbands) or additionally the binarized image, which in total produce a final image that has four channels or more. In the case of multiple networks, multiple CNNs are considered, wherein the input image for each stream includes the original image and one of the subband images or the binarized image so that the final image for each stream has two channels. The segmentation result of each CNN is integrated to produce the final segmentation map.

In the first method (Method a), the original image is decomposed into wavelet subbands and the original image is binarized by the structural symmetric pixels (SSPs) method (single network). In the second method (Method b), an implementation of ‘Method a’ with multiple networks is considered. In the third method (Method c), fewer number of channels are used for reducing the computational cost.

6) **University of Sousse, ISITCom, Tunisia and University of Sfax, CRNS, Tunisia** (*Mohamed Ali Souibgui, Yousri Kessentini*)

In this method, document image binarization is considered as an image to image translation task where the goal is to generate binary document images given the degraded ones. In view of this, a conditional generative adversarial network is used, called DE-GAN (for Document Enhancement conditional Generative Adversarial Network). GANs usually contain two networks, a generator and a discriminator. To train the model, patches of size 256x256 are used taken from the previous DIBCO datasets. After training, the generator could be used for binarizing the new document images.

7) **Universitas Syiah Kuala, Indonesia** (*Khairun Saddami*)

Method a:
The first method is the combination of Niblack and Wolf (CNW) binarization method that actually combined Niblack and Wolf thresholding formula. It has been presented in [12].
Method b:
The second method is the combination of a Local and a Global (CLG) binarization method that actually combines a local adaptive and global thresholding formula. A detailed decsription is given in [13].

8) **Lenovo Research, Beijing, P.R, China** *(Wanjun Lyu, Hui Li, Luyan Wang, Xiaoping Zhang, Yaqiang Wu)*

In this method, two segmentation networks have been used to implement image binarization. The training dataset originates from previous DIBCO competitions, and contains 116 images labeled at the pixel level. First, we apply two binarization methods on the input color image, one is Howe's binarization method, and the other is a combination of improved guided filter and Bilateral filter. Then, combine grayscale image with the previous two binarization results as input of the segment network. Both of the submitted methods rely upon an encoder-decoder structure. A modified Resnet101 is used as the backbone of the encoder network. The modification is replacing the first 7*7 convolution layer with two 3*3 convolution layers. There are two differences between the two distinct submissions:
Method a: In the encoder structure, the last three convolution layers are replaced by dilation convolution layers.
Method b: In the decoder structure, the first submission is mainly based on deeplab-v3-plus network, and the last 4X interpolation layer is replaced by two 2X interpolation layers. The submission is mainly based on U-Net.
Finally, a post-processing method is used to preserve text stroke connectivity.

9) **Jadavpur University, Kolkata, India and University at Buffalo, USA** (*Showmik Bhowmik, Ram Sarkar and David Doermann*)

The submitted binarization method called GiB [14] is a game theory inspired method having two key modules - background separation and binarization. At first from a grayscale image we approximate and remove the possible background information using a customized 'inpainting' method [15]. In this way, some noisy pixels also get eliminated, and we get a background separated image. After that, we design a two-player game, and implement the same at the pixel level. We scan an overlapping window over, and for each window we consider the central pixel as the first player and rest of the eight pixels together are considered as the second player. We compute the Nash equilibrium for each such game. Payoff value for the central pixel, at Nash equilibrium state, is taken as a feature for each game. Besides, we compute two more features - the central pixel intensity and intensity difference between central pixel and the pixel having maximum intensity among its 8 neighbors. Based on these three features, all the pixels are grouped into three clusters. We use K-means clustering algorithm, with dynamically selected initial cluster centers, to group the pixels. The cluster with lowest variance is considered as background. We take the ratio between the variances of remaining two unlabeled clusters. If this ratio is less than a threshold then it indicates these are similar, thus merged. Otherwise, we combine the cluster having less variance with previously identified background cluster.

10) **Jadavpur University, Kolkata, India**

Method a – authored by (*Arpan Basu, Riktim Mondal, Manosij Ghosh, Vivek Roy, Showmik Bhowmik, Ram Sarkar*)

In this method, a U-NET based convolutional network with skip-connections between the down-sampling and up-sampling layers is used. The image is first down-sampled using convolutional blocks with a stride of 2, and then up-sampled using transposed convolutional blocks to the original dimension. The input to this model consists of 256 x 256 patches of the original images. This means an input is a 256 x 256 x 3 tensor. The output produced by the model is a 256 x 256 x 1 tensor with each value between 0 and 1. This is concatenated along the last dimension to produce a 256 x 256 x 3 greyscale patch. The model is trained using the Adam optimization algorithm using binary cross-entropy between each input patch and the corresponding ground-truth patch as the loss. The image is then thresholded at 0.5 to produce a binarized patch. To produce output, we take 256 x 256 overlapping patches across the image, suitably padded with edge pixels if required, with a stride of 64 along both horizontal and vertical directions.

Method b – authored by (*Soulib Ghosh, Suman Kumar Bera, Showmik Bhowmik, Ram Sarkar*)

This method uses an ensemble of three clustering algorithms, namely Fuzzy C-Means, K-Medoids and K-Means++ for document image binarization. Before this, it performs a noise reduction step on the input grayscale image to eliminate the background variation. In this step, initially the background of the input image is estimated using an inpainting method [15], then the estimated background image is used to normalize the input image so that the influence of high and very low frequency noise can be suppressed. This background suppressed image is then used as an input to the binarization step. In this step, pixels of the background suppressed image are clustered into two groups, foreground and background, using Fuzzy C-Means, K-Medoids and K-Means++ algorithms. The final decision relies upon a voting process. In this process, at first, the quality of the clusters generated by these three algorithms is measured separately using the Davies-Bouldin (DB) cluster validity index [16]. The algorithm with the smallest DB index value is considered as the "decider". Then, for each pixel is checked how the other two algorithms have labeled it. If there is a conflict, then the label provided by the decider is considered as the final label of this pixel. Otherwise, the common label is considered as the final one. Finally, a post-processing technique is applied to preserve the stroke connectivity so that the quality of text regions gets improved.

11) **University of Alicante, Spain** (*Jorge Calvo-Zaragoza and Antonio Javier Gallego*)

From a machine learning point of view, image binarization can be formulated as a two-class classification task at pixel level. The presented strategy follows this idea and, therefore, basically consists in learning which label must be given to every single pixel of the image. Since we are dealing with images of documents, the set of labels are defined as foreground and background. Specifically, we make use of Convolutional Neural Networks. These networks involve multi-layer architectures that perform a series of transformations to the input signal. The parameters of these transformations are adjusted through a training process. In this case, our proposal is to consider an image-to-image convolutional architecture, which is trained to convert an input image into its binarized version. This has a number of advantages such that the classification of each pixel of the image is not produced independently, but also takes into account the label to be assigned to its neighbors. In addition, several pixels can be processed at the same time, thereby leading to higher efficiency than a pixel-wise classification approach. More importantly, the method can be straightforwardly adapted to many different scenarios or graphical domains by just changing the training data.
Once the network has been properly trained an image can be parsed through the network, after which a selection level is assigned to each input pixel. In practice, the network hardly outputs either 0 or 1 but an intermediate value. Therefore, a thresholding process is still necessary to convert the obtained scores into actual binary values. Those pixels whose selection value exceeds a certain threshold are considered to belong to the foreground of the document image, whereas the others are labeled as background. More details about the submitted method can be found in [17].

12) **Larbi Tebessi University, Tébessa, Algeria** (*Abdeljalil Gattal*)

This method [18] for Handwritten Document Image Binarization based on K-Means algorithm can be performed mainly in two steps: Preprocessing and clustering. The preprocessing step is included all the functions necessary to produce the best grayscale image. Next, the K-Means clustering algorithm allows setting the image pixels into the corresponding foreground, background or noise cluster.

13) **The Australian National University, Canberra, Australia** (*Hanif Rasyidi*)

In this competition, a pattern-based binarization method is proposed that is called HandwritteNet. This model used a fully convolutional network [19] to analyze the text pattern on the document, then apply a pixel-based semantic segmentation to produce a binary image that contains text. We train the model using the available DIBCO dataset available from this competition's website, as well as Nabuco dataset that is available on the DIB portal (https://dib.cin.ufpe.br/). The proposed model contains three different part: feature extraction backbone, feature merging, and the final output layer for pixel segmentation. This idea is based on EAST model [20] which uses different backbone and output layer to detect text in the scene images. We use ResNet50 model as a backbone [21] due to the implementation of residual connection that helps to prevent the lost of low level information (Method a). We also made a variation of HandwritteNet called HandwritteNet-Mobile, that uses a less-costly MobileNetV2 [22] as the backbone (Method b). Compared to the ResNet101, a deeper version of ResNet50 model, MobileNetV2 only uses parameters with the number less than 10 percent of ResNet101 to doing the similar task. That efficiency helps the model to run faster, even though the model is harder to train. Both of the backbone model is pre-trained using ImageNet dataset [23] with 1000 class before connected to our model.

14) **Universidade de São Paulo, São Paulo, Brazil** (*Nury Yuleny Arosquipa Yanque, Gustavo Enrique Salazar Torres, Roberto Hirata Junior*)

The method relies upon a supervised machine learning technique. The feature vectors are composed by a combination of the following: (i) Binary output values from state-of-art methods like Otsu, Niblack, Sauvola, Su and Howe; (ii) Binary image output from GridLSTM proposed by Westphal; (iii) Family of texture features called 'Relative Darkness Index' proposed by Wu and (iv) the grayscale intensity value of original image.

These vectors are extracted for every pixel of the gray-scaled original image. The dataset composed by these vectors and the foreground/background labels is used to train an XGBoost classifier that predicts if the pixel belongs to a text or background region. The output image is post-processed by a morphological operation in order to improve the quality of the image.

15) **Istanbul Technical University, Turkey** (*Yasin YILDIRIM*)

The submitted document image binarization method consists of 3 main steps which are preprocessing, optimization and thresholding. At the preprocessing stage the input image is converted to grayscale and then a 9x9 adaptive Wiener filter is applied to reduce noise. At the second step, an optimization is used that aims to improve contrast and brightness of the document image for successful thresholding. Finally, Otsu thresholding is applied as a final step in order to binarize document image.

III. EVALUATION MEASURES

For the evaluation, the measures used comprise an ensemble of measures that are suitable for evaluation purposes in the context of document image analysis and recognition. These measures consist of (i) F-Measure (*FM*), (ii) pseudo-FMeasure (F_{ps}), (iii) *PSNR* and (iv) Distance Reciprocal Distortion (*DRD*).

A. F-Measure

$$FM = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \qquad (1)$$

where $\text{Recall} = \frac{TP}{TP+FN}$, $\text{Precision} = \frac{TP}{TP+FP}$

TP, FP, FN denote the True Positive, False Positive and False Negative values, respectively.

B. pseudo-FMeasure

Pseudo-FMeasure $\boldsymbol{F_{ps}}$ is introduced in [24] and it uses pseudo-Recall $\boldsymbol{R_{ps}}$ and pseudo-Precision $\boldsymbol{P_{ps}}$ (following the same formula as F-Measure). The pseudo Recall/Precision metrics use distance weights with respect to the contour of the ground-truth (GT) characters. In the case of pseudo-Recall, the weights of the GT foreground are normalized according to the local stroke width. Generally, those weights are delimited between [0,1]. In the case of pseudo-Precision, the weights are constrained within an area that expands to the GT background taking into account the stroke width of the nearest GT component. Inside this area, the weights are greater than one (generally delimited between (1,2]) while outside this area they are equal to one.

C. PSNR

$$PSNR = 10\log\left(\frac{C^2}{MSE}\right) \qquad (2)$$

where $MSE = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}(I(x,y)-I'(x,y))^2}{MN}$

PSNR is a measure of how close is an image to another. The higher the value of *PSNR*, the higher the similarity of the two images. Note that the difference between foreground and background equals to *C*.

D. Distance Reciprocal Distortion Metric (DRD)

The Distance Reciprocal Distortion Metric (*DRD*) has been used to measure the visual distortion in binary document images [25]. It properly correlates with the human visual perception and it measures the distortion for all the $\boldsymbol{S}$ flipped pixels as follows:

$$DRD = \frac{\sum_{k=1}^{S} DRD_k}{NUBN} \qquad (3)$$

where NUBN is the number of the non-uniform (not all black or white pixels) 8x8 blocks in the GT image, and DRD_k is the distortion of the *k*-th flipped pixel that is calculated using a 5x5 normalized weight matrix $\boldsymbol{W_{Nm}}$ as defined in [25]. DRD_k equals to the weighted sum of the pixels in the 5x5 block of the $\boldsymbol{GT}$ that differ from the centered k^{th} flipped pixel at (x,y) in the binarization result image $\boldsymbol{B}$ (Eq. 4).

$$DRD_k = \sum_{i=-2}^{2}\sum_{j=-2}^{2} |GT_k(i,j) - B_k(x,y)| \times W_{Nm}(i,j) \qquad (4)$$

IV. EXPERIMENTAL RESULTS

The DIBCO 2019 testing dataset consists of 20 document images that can be divided into two distinct categories:

- CATEGORY I: A collection of 10 historical machine-printed and handwritten document images for which representative degradations appear. The handwritten document images originate from collections that belong to READ project [26], while the machine-printed documents of the dataset originate from collections which are dated in the beginning of 19^{th} century and have been printed in several places including Venice, Leipzig and Constantinople.

- CATEGORY II: It concerns a papyri dataset (10 images) that has been chosen to reflect the diversity of material ancient historians are confronted with. The papyri all bear a literary text but they come from different places in Egypt, were written at various periods of Antiquity (3rd c. BCE – 6th c. CE) using different kinds of papyrus quality, ink and handwriting styles. They show various states of preservation (holes, erased ink) and restoration (dust, misplaced fibers). The images were provided by the institutions that own the

papyri, therefore they do not either have homogeneous properties (different digitization material, resolution, lighting, etc). Further details about these papyri are given in [27].
For all document images the associated ground truth was built manually for the evaluation. All testing images of Category I and Category II are shown in Fig. 1(a) and Fig. 2(a), respectively.

The evaluation was based upon the four distinct measures presented in Section III. The detailed evaluation results along with the final ranking are shown in Tables I-III. The final Ranking was calculated after first, sorting the accumulated ranking value for all measures for each test image. The summation of all accumulated ranking values for all test images denote the final score which is shown in Table I at column "Total Score". Additionally, the evaluation results for the widely used binarization techniques of Otsu [6] and Sauvola [28] are also presented. At Table II-III we provide the performance of each algorithm for the portion of the testing dataset that belongs to Category I and II, respectively. Overall, the best performance is achieved by ***Method 10b*** which has been submitted by ***Soulib Ghosh, Suman Kumar Bera, Showmik Bhowmik, Ram Sarkar*** affiliated to **Jadavpur University, Kolkata, India**. The binarization results of this algorithm for each image of the testing dataset is shown in Fig. 1(b) and Fig. 2(b) corresponding to Category I and II, respectively.

V. CONCLUSIONS

The major goal which has been achieved via the DIBCO competition series since the beginning of this initiative is the better understanding on the algorithmic behavior of document image binarisation methods in an objective evaluation context. Taking into account the participating methods and their performance, several conclusions are drawn that could be used as valuable hints for the research community working on improving document image binarization methods.

A critical difference of this year's competition compared to those of previous years is the dataset selection which had in mind to present a challenging collection with no similarities with the already available DIBCO dataset collection. This played a crucial role to the performance of supervised approaches that were trained with the existing DIBCO collections. In particular, many supervised approaches, including some current state of the art that have been published the last two years did not succeed to show the expected high performance. Contrary to this, methods that were based in data independent methodologies, like locally adaptive methods as Niblack [29] have demonstrated a superior performance. It is worth noting that in any case the overall performance did not reach high levels due to also the challenging datasets.

As has been described in Section IV, we encountered document images that have been categorized in two types. The dataset of Category I comprises document images that reflect difficult cases including combination of handwritten and machine-printed text in a single document while the dataset of Category II comprises papyri images that appear for the first time in a DIBCO competition. The deviation of both categories in relation to the previous years' DIBCO collection had major impact in the methodology that had the best performance. However, it is not so important to stick on the single formal winner since the difference in performance with the methods that ranked in the next positions is not high enough to dither their importance.

Most of the submitted methods rely in various deep learning architectures that addressed the problem of document image binarisation either as a segmentation or as a classification problem. The network that was chosen more frequently from the participating research teams was the U-Net which appeared in four distinct variations.

As in previous years' DIBCO challenges, standard approaches which apart of the locally adaptive Niblack [29], also the global Otsu algorithm [6] and the method of Howe [30] was used in certain proposed approaches. Last but not least, it has been a general rule the use of an explicit post-processing stage which in certain cases is coupled with the use of a pre-processing stage, proving that those stages have a major impact on the success of the binarization process.

TABLE I. OVERALL EVALUATION RESULTS FOR ALL METHODS SUBMITTED TO DIBCO 2019.

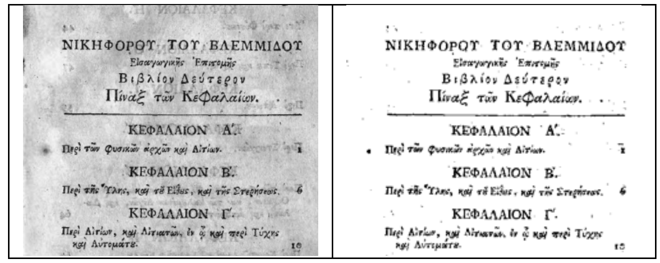
Overall Rank	Method	Total Score	FM	F_{ps}	PSNR	DRD
1	10b	645	72.875	72.15	14.475	16.235
2	9	701	71.625	70.78	14.15	16.71
3	2a	741	70.43	69.84	15.31	8.05
4	15	801	68.42	68.44	13.905	18.25
5	3b	877	62.985	61.01	14.32	10.84
6	2b	882	62.705	64.2	14.79	9.14
7	3a	883	63.48	61.515	14.135	11.475
8	7a	942	56.24	53.96	10.92	67.17
9	4	954	56.31	53.03	10.63	84.66
10	12	994	60.145	56.7	11.745	36.52
11	5c	1003	58.16	55.815	12.325	20.86
12	10a	1007	60.145	57.625	12.415	18.275
13	1b	1012	59.25	56.51	12.71	17.54
14	5b	1035	56.34	54.205	12.435	20.255
15	5a	1036	57.975	56.295	13.085	16.695
16	7b	1050	57.345	55.23	11.07	30.09
17	6	1081	55.975	53.44	12.29	16.575
18	14	1128	53.715	53.76	12.825	20.19
19	1a	1137	55.415	52.76	12.205	21.6
20	8a	1138	46.63	44.06	13.09	15.57
21	13a	1142	54.415	50.68	12.04	17.35
22	11	1200	51.605	50.03	11.765	22.925
23	8b	1278	40.67	39.655	13.015	14.22
24	13b	1295	46.695	41.29	9.345	40.95

TABLE II. DETAILED EVALUATION RESULTS FOR ALL METHODS SUBMITTED TO DIBCO 2019 FOR DATASET OF CATEGORY I

Rank	Method	Score	FM	F_{ps}	PSNR	DRD
1	10b	288	77.76	76.42	16.81	5.60
2	9	307	76.09	75.04	16.46	6.88
3	7a	331	74.62	72.78	16.00	8.00
4	15	358	72.92	73.05	16.43	4.89
5	4	382	71.76	66.81	14.83	7.60
6	2a	445	64.45	62.02	15.35	4.94
7	5c	464	63.54	58.99	13.32	13.01
8	5b	468	63.76	59.04	13.43	12.72
9	3b	472	62.65	58.98	15.23	4.95
10	12	483	65.02	60.87	14.04	9.62
11	3a	489	62.10	58.48	15.05	5.35
12	5a	491	64.78	60.30	14.38	7.85
13	8a	497	66.86	60.18	14.32	7.88
14	2b	503	59.24	59.32	15.13	4.79
15	7b	513	60.71	57.11	12.34	16.39
16	6	569	56.25	52.34	12.66	8.27
17	10a	573	57.72	53.40	13.03	11.68
18	1a	586	58.04	53.15	12.76	11.61
19	1b	593	55.75	50.61	12.66	12.18
20	8b	596	57.03	54.08	14.12	6.78
21	11	598	54.93	51.70	12.77	13.28
22	14	602	52.24	55.03	14.72	5.20
23	13a	674	47.80	41.45	12.34	10.71
24	13b	680	46.76	39.55	10.45	20.78
-	Otsu	-	72.33	70.44	15.37	9.55
-	Sauvola	-	61.7	58.78	12.64	15.44

TABLE III. DETAILED EVALUATION RESULTS FOR ALL METHODS SUBMITTED TO DIBCO 2019 FOR DATASET OF CATEGORY II

Rank	Method	Score	FM	F_{ps}	PSNR	DRD
1	2a	296	76.41	77.66	15.27	11.16
2	10b	357	67.99	67.88	12.14	26.87
3	2b	379	66.17	69.08	14.45	13.49
4	9	394	67.16	66.52	11.84	26.54
5	3a	394	64.86	64.55	13.22	17.6
6	3b	405	63.32	63.04	13.41	16.73
7	1b	419	62.75	62.41	12.76	22.9
8	10a	434	62.57	61.85	11.8	24.87
9	15	443	63.92	63.83	11.38	31.61
10	13a	468	61.03	59.91	11.74	23.99
11	12	511	55.27	52.53	9.45	63.42
12	6	512	55.7	54.54	11.92	24.88
13	14	526	55.19	52.49	10.93	35.18
14	7b	537	53.98	53.35	9.8	43.79
15	5c	539	52.78	52.64	11.33	28.71
16	5a	545	51.17	52.29	11.79	25.54
17	1a	551	52.79	52.37	11.65	31.59
18	5b	567	48.92	49.37	11.44	27.79
19	4	572	40.86	39.25	6.43	161.72
20	11	602	48.28	48.36	10.76	32.57
21	7a	611	37.86	35.14	5.84	126.34
22	13b	615	46.63	43.03	8.24	61.12
23	8a	641	26.4	27.94	11.86	23.26
24	8b	682	24.31	25.23	11.91	21.66
-	Otsu	-	33.27	34.67	9.9	38.93
-	Sauvola	-	23.33	20.74	2.78	209.36



(a) (b)

Fig. 1. (a) The DIBCO 2019 testing dataset of Category I (b) Binarization results from the winner algorithm of DIBCO 2019.

(a) (b)

Fig. 2. (a) The DIBCO 2019 testing dataset of Category II (b) Binarization results from the winner algorithm of DIBCO 2019.

ACKNOWLEDGEMENT

This research has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH-CREATE-INNOVATE (project code: T1EDK-01939)

References

[1] B. Gatos, K. Ntirogiannis and I. Pratikakis, "ICDAR 2009 Document Image Binarization Contest (DIBCO2009)", 10th International Conference on Document analysis and Recognition (ICDAR'09), July. 26-29, 2009, Barcelona, Spain, pp. 1375-1382.

[2] I. Pratikakis, K. Zagoris, G. Kaddas and B. Gatos, "ICFHR 2018 Competition on Handwritten Document Image Binarization (H-DIBCO 2018)", 16th International Conference on Frontiers in Handwriting Recognition (ICFHR'18), August 5-8, 2018, Niagara Falls, NY, USA, pp. 619-623, 2018.

[3] I. Pratikakis, K. Zagoris, G. Barlas and B. Gatos, "ICDAR 2017 Competition on Document Image Binarization (DIBCO 2017)", 14th IAPR International Conference on Document Analysis and Recognition (ICDAR'17), November 09-15, 2017, Kyoto, Japan, pp. 1395-1403, 2017.

[4] B. Epshtein, E. Ofek, Y. Wexler, "Detecting text in natural scenes with stroke width transform," in Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, 2010, pp. 2963-2970. doi: 10.1109/cvpr.2010.5540041.

[5] Kong X., Sun G., Wu Q., et al. "Hybrid pyramid U-Net model for brain tumor segmentation," in Proceedings of the 10th IFIP TC-12 International Conference on Intelligent Information Processing (IIP 2018), Nanning, China, 2018, pp. 346-355.

[6] N. Otsu, "A threshold selection method from gray-level histograms", IEEE Trans. Sys., Man., Cyber., vol. 9, 1979, p. 62–66.

[7] L. Zhou, C. Zhang, and M. Wu, "D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction," in Proceedings of the 2018 IEEE/CVF Conference on CVPR Workshops (CVPRW), Salt Lake City, UT, 2018, pp. 192-1924.

[8] J. Zhuang, "Laddernet: Multi-path networks based on unet for medical image segmentation," ArXiv 2018. https://github.com/juntang-zhuang/LadderNet

[9] S. He, L. Schomaker, "DeepOtsu: Document Enhancement and Binarization using Iterative Deep Learning", Pattern recognition, vol. 91, pp. 379-390, 2019.

[10] Jia, Fuxi, Cunzhao Shi, Kun He, Chunheng Wang, and Baihua Xiao. "Degraded document image binarization using structural symmetry of strokes." Pattern Recognition 74 (2018): 225-240.

[11] Younes Akbari, Alceu S. Britto~Jr., Somaya Al-Maadeed, Luiz S. Oliveira, "Binarization of Degraded Document Images using Convolutional Neural Networks based on predicted Two-Channel Images", Accepted in ICDAR 2019.

[12] K. Saddami, K., Afrah, P., Mutiawani, V., & Arnia, F. "A New Adaptive Threshoding Technique for Binarizing Ancient Document". In International Conference of INAPR, 2018 (Vol. 1, pp. 57-61). IEEE

[13] S. Saddami, K., Munadi, K., Away, Y., & Arnia, F. (2019). "Combination Local and Global Thresholding Method for Binarizing Ancient Jawi Document". Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK). In Press

[14] S. Bhowmik, R. Sarkar, B. Das, and D. Doermann, "GiB: a Game theory Inspired Binarization technique for degraded document images," IEEE Transactions on Image Processing, vol. 28, no. 3, pp. 1443–1455, 2019

[15] K. Ntirogiannis, B. Gatos, and I. Pratikakis, "A combined approach for the binarization of handwritten document images," Pattern Recognition Letters, vol. 35, pp. 3–15, 2014.

[16] D.L. Davies and D.W. Bouldin, "A cluster separation measure", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 1, Issue 2, 224-227, 1979.

[17] M Jorge Calvo-Zaragoza, Antonio-Javier Gallego: A selectional auto-encoder approach for document image binarization. Pattern Recognition 86: 37-47, 2019.

[18] A. Gattal, F. Abbas, M. R. Laouar, Automatic Parameter Tuning of K-Means Algorithm for Document Binarization, The 7th International Conference on Software Engineering and New Technologies "ICSENT'2018", Hammamet, Tunisia.

[19] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," In CVPR 2015, pp. 3431–3440.

[20] X. Zhou, C. Yao, H. Wen, Y. Wang, S. Zhou, W. He, and J. Liang, "East: an efficient and accurate scene text detector," in Proceedings of CVPR, 2017, pp. 5551–5560.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of CVPR, June 2016.

[22] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4510–4520.

[23] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in 2009 IEEE conference on computer vision and pattern recognition, 2009, pp. 248–255.

[24] K. Ntirogiannis, B. Gatos and I. Pratikakis, "Performance Evaluation Methodology for Historical Document Image Binarization", IEEE Transactions on Image Processing, vol. 22, no.2, pp. 595-609, 2013.

[25] H. Lu, A. C. Kot and Y.Q. Shi, "Distance-Reciprocal Distortion Measure for Binary Document Images", IEEE Signal Processing Letters, vol. 11, No. 2, pp. 228-231, 2004.

[26] READ project (http://read.transkribus.eu/)

[27] https://d-scribes.philhist.unibas.ch/en/case-studies/iliad

[28] J. Sauvola, M. Pietikainen, "Adaptive document image binarization", Pattern Recognition, Vol. 33, No. 2, pp. 225-236, 2000.

[29] W. Niblack, "An introduction to Digital Image Processing", Strandberg Publishing Company, Birkeroed, Denmark, 1985

[30] N. Howe, "Document Binarization with Automatic Parameter Tuning", International Journal on Document Analysis and Recognition, vol. 16, no. 3, pp. 247-258, 2013.