

# Cluster-based Sample Selection for Document Image Binarization

Amandus Krantz  
Department of Computer Science  
Blekinge Institute of Technology  
Karlskrona, Sweden  
krantz.amandus@gmail.com

Florian Westphal  
Department of Computer Science  
Blekinge Institute of Technology  
Karlskrona, Sweden  
florian.westphal@bth.se

**Abstract**—The current state-of-the-art, in terms of performance, for solving document image binarization is training artificial neural networks on pre-labelled ground truth data. As such, it faces the same issues as other, more conventional, classification problems; requiring a large amount of training data. However, unlike those conventional classification problems, document image binarization involves having to either manually craft or estimate the binarized ground truth data, which can be error-prone and time-consuming. This is where sample selection, the act of selecting training samples based on some method or metric, might help. By reducing the size of the training dataset in such a way that the binarization performance is not impacted, the required time spent creating the ground truth is also reduced. This paper proposes a cluster-based sample selection method that uses image similarity metrics and the relative neighbourhood graph to reduce the underlying redundancy of the dataset. The method, implemented with affinity propagation and the structural similarity index, reduces the training dataset on average by 49.57% while reducing the binarization performance only by 0.55%.

**Keywords**—document image binarization; sample selection; neural networks; computer vision

## I. INTRODUCTION

An important initial step when performing computer vision aided analysis of physical documents is document image binarization; the act of separating text from its background. Performing document image binarization not only makes the text more readable by removing degradations, but the high contrast result can also be easier for optical character recognition (OCR) engines to read. One of the more common and often best performing methods for document image binarization is to train a neural network on datasets containing images of physical documents and their binarized ground truth data. The issue with this is that creating the ground truth data, either manually or through estimation, can be very time consuming and sometimes prone to errors, especially when using neural networks which are known to require vast amounts of labelled training data. Creating such data for image binarization is a time consuming and error-prone process, which indicates a need for a method, which can reduce the amount of training data required for an acceptable performance. Data augmentation may be used to decrease the amount of data needed, however, it still leaves

the question of what data to augment and train on. This is where sample selection can help.

Rather than selecting a random portion of the dataset to use for training, sample selection selects specific samples from the dataset. The selected training samples should be representative of the dataset as a whole and should not reduce the testing performance by a significant amount when trained on. The most common way of selecting these samples is to look at the decision boundary of the problem and remove as many samples as possible without distorting the boundary [1], [2], [3].

In this paper, we design and test a sample selection method specifically for document image binarization with the aim of reducing the amount of training data required for acceptable performance. This, in turn, would also reduce the number of man-hours required to label such data, without a significant reduction of the binarization performance. The proposed method makes use of affinity propagation to find “natural” clusters based on image similarity. These clusters are then used together with a relative neighbourhood graph to find and extract samples that lie close to the cluster boundaries, which are then used for training.

The contributions of this paper are the following:

- 1) A sample selection method specifically designed for document image binarization is proposed, which does not rely on available ground truth data to work.
- 2) The sample selection method is shown through experimentation on the DIBCO and H-DIBCO datasets to reduce the training dataset size by as much as 49.57% on average, though with a 0.55% decrease in median binarization performance.

## II. RELATED WORK

The problem with many neural networks, especially those dealing with computer vision, is the large amount of data and resulting long training times that are required to get acceptable testing results. This issue has prompted many researchers to look at the impact of redundancy in datasets and how redundant samples may be removed to reduce the training times and required data. One of the earliest attempts at this was presented in 1968 by Hart [1] in the context of nearest neighbour classification. It was proposed that rather

than using the entire dataset, one should focus on samples that exist close to class boundaries. The theory was that these samples would be located further away from the bulk of the samples in the class, and should thus have more information about where the class boundary should be drawn.

Hart’s method was later adapted by Bhattacharya *et al.* [4] into a method that uses graph theory, specifically the Gabriel graph and relative neighbourhood graph (RNG), to approximate the decision boundary and then select samples based on this approximation. The method works by constructing an undirected graph out of the dataset and then marking every node in the graph that has neighbouring nodes of a different class. The decision boundary is estimated to lie between these marked nodes, often referred to as bridge vectors. This method results in a very good approximation of the decision boundary. Using Hart’s idea, that samples existing close to the decision boundary are what matter the most, every unmarked sample is then removed, resulting in a large reduction of the training dataset but with an intact decision boundary.

This method of using graphs to estimate the decision boundary of the dataset has been tested and implemented several times since. Goto *et al.* [5] use the RNG to reduce a training dataset consisting of handwritten digits to 10% of its original size. When a support vector machine (SVM) was trained on the reduced dataset, it achieved similar results as an SVM trained on the complete dataset, and better results than an SVM trained on a random subset of the same size. Rayar *et al.* [2] achieved similar results using the RNG to reduce a handwritten digit dataset by as much as 76% without any significant degradation in classification performance when tested with a convolutional neural network.

In contrast to these approaches to sample selection, our proposed approach does not require the presence of ground truth data, which allows it to reduce the human labelling effort. Furthermore, our approach is designed to address sample selection for image binarization in particular, which is different from the classification problems addressed by the described approaches. This is the case since a selected training sample in binarization has several class labels, each pixel in a training patch has its own label, while in other classification problems there is only one label per sample.

### III. BACKGROUND

#### A. U-Net

In an attempt to reduce the amount of data required for bio-medical image segmentation (a field closely related to document image binarization), Ronneberger *et al.* [6] created the U-Net; a fully convolutional neural network that reduces the amount of required training data through a combination of downsampling, upsampling, and data augmentation. U-Nets use contracting and expanding paths to first create a

low-resolution representation of the input image. This low-resolution representation is then used to extrapolate a high-resolution image that should only contain the important parts of the input image.

The 2017 edition of the Document Image Binarization Competition (DIBCO) [7] showed that the U-Net is able to outperform many other document image binarization methods. Due to this high performance, the U-Net was chosen to be used as binarization algorithm in this study.

#### B. The Structural Similarity Index

The structural similarity (SSIM) [8] is used to measure the similarity between samples in the dataset. Originally meant to measure the quality of digital image compression, the SSIM index compares the luminance, contrast, and structural information of two images and assigns the image pair a similarity score between 0.0 and 1.0, where 0.0 means the images are completely different and 1.0 means the images are identical. The SSIM index strikes a good balance of being computationally light, while still being a very robust metric of similarity when compared to other metrics, such as the Euclidean distance which does not handle transformations, such as rotation and translation.

#### C. Affinity Propagation

The clustering method of choice for this study is affinity propagation [9], [10], [11], a clustering method that creates clusters based on representative nodes. Each node in the dataset selects a representative node based on their similarity and a predefined preference. In this case the preference was set to the median of the input similarities for the dataset. When a node has selected a representative node, it is added to that node’s cluster. This is repeated either until convergence or until a predefined number of iterations is reached. Affinity propagation has the advantage of not requiring the number of clusters present in the dataset to be known.

### IV. CLUSTER BASED SAMPLE SELECTION

This section describes the proposed method for implementing sample selection for document image binarization. The method uses the following steps to select samples that lie close to the boundaries of a cluster:

- 1) Create clusters based on a similarity measure
- 2) Build relative neighbourhood graph based on a distance metric
- 3) Use clusters and relative neighbourhood graph to find bridge vectors

#### A. Creating the Clusters

Previous papers on sample selection [1], [2], [3], [4], [5] use classes to, based on some method or metric, estimate where the decision boundary lies and use this information to select training samples. However, training samples in

document image binarization cannot be represented by a single class. It is thus proposed to use unsupervised clustering methods to approximate single class representations based on the similarity between the training samples. The hypothesis is that images that look similar to each other should logically contain roughly the same information and should thus teach the learner similar things, just like in a classification problem where samples with similar features are likely to be part of the same class. In this study, affinity propagation was used to find clusters based on the image similarity scores calculated by the SSIM index.

### B. Building the Relative Neighbourhood Graph

With the clusters working as stand-ins for classes, the next step is to select samples. Based on the work of Bhattacharya *et al.* [4], an RNG is created out of the entire dataset. The definition of a neighbour in the RNG is “two points that are at least as close to each other as they are to any other point” [12], which indicates that a distance metric of some sort is required to create neighbours. Rayar *et al.* [2] and Goto *et al.* [5] who are also working with images, though in the context of handwritten digit classification, use the Euclidean distance between images. The issue with this is, as mentioned earlier, that the Euclidean distance is quite naive and does not take into consideration things like rotation and translation. Some metrics attempt to address this [13], however, this method already includes a measure that can be used instead; the similarity metric used for clustering. Similarity metrics and distance metrics essentially measure the same thing; the difference between specific features of two entities. Where a similarity metric measures the difference in similarity, a distance metric measures the difference in location. Using a similarity metric in the RNG requires a slight change to the definition of neighbours to the following: two points are neighbours if they are at least as similar to each other as they are to any other point. For completeness and comparison with previous work, the Euclidean distance will also be used to create an RNG.

### C. Finding the Bridge Vectors

The actual sample selection is performed once both the clusters and the RNG have been created. It is done by going through each node in the RNG and comparing the cluster labels of its neighbours with the cluster label of the node itself. If the labels do not match, both nodes are added to a list of bridge vectors that keep track of nodes that “bridge” neighbourhoods. The cluster boundary is estimated to run exactly in the middle of a bridge vector (see Figure 1), which means that the nodes in the bridge vectors should be the nodes that lie the closest to the cluster boundary. These nodes should be the samples that are the least similar to the other samples in the cluster, reducing their redundancy.

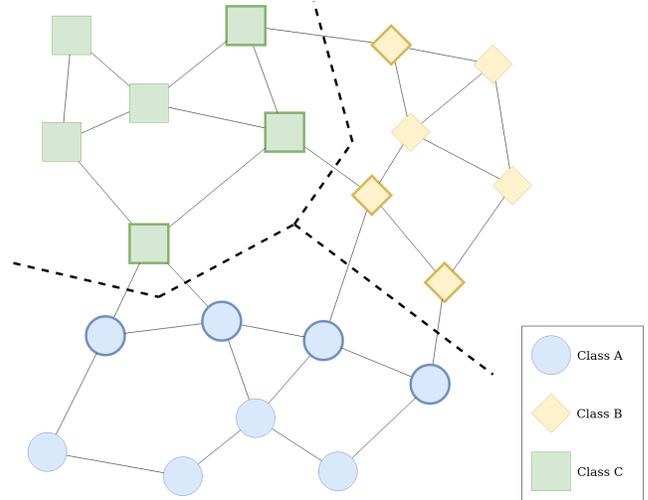


Figure 1. Example of bridge vectors on toy data. Bridge vectors are highlighted with thicker borders. Dashed line represents estimated decision boundary.

## V. EXPERIMENT DESIGN

All experiments were run on a computer with an Intel i9-7900X CPU @ 3.3 GHz, 32 GB DDR4 RAM and two Nvidia GeForce GTX 1080 Ti. The U-Net [6] implementation robin<sup>1</sup>, written with Keras 2.2.4 [14], was used to create and train U-Nets for the experiments.

The sample selection methods were implemented in Python with Scikit-learn [15] for the clustering and Networkx [16] for the RNG<sup>2</sup>.

### A. Datasets

The datasets used for training were compiled from the publicly available datasets used in the bi-annual Document Image Binarization Contest (DIBCO) [17], [18], [19], [20] and Handwritten Document Image Binarization Contest (H-DIBCO) [21], [22], [23], [24], [25]. The dataset from each year was used for testing once, with the rest of the years being used for training, resulting in nine test datasets and corresponding training datasets. These datasets contain document images, both handwritten and printed, and are meant to be representative of the kinds of images that might be encountered in a real document image binarization application.

Each image in each training dataset and its accompanying ground truth image was split into  $128 \times 128$  grayscale images. In the cases where the  $128 \times 128$  sized image patch exceeded the boundaries of the original image, it was padded with solid white. Each of the sample selection methods was then run on each of the training datasets consisting of the

<sup>1</sup><https://github.com/masyagin1998/robin>

<sup>2</sup>The implementation, as well as the raw data are available at <https://github.com/krrntz/Cluster-based-Sample-Selection>

created patches. This results in three training datasets per testing dataset; two sample selected, and one non-sample selected. The non-sample selected dataset was used as a control to check if the sample selection methods result in any significant reduction of binarization performance.

Data augmentation has been intentionally disabled for all training. This is to make sure that there is no “smoothing” of the performance, which would make the number and nature of the selected samples matter less.

### B. Training Setup

For each of the created training datasets, a U-Net was trained from scratch. Early stopping was used during this training to prevent overfitting. This causes training to stop if there have been no changes in validation performance after a specified number of epochs. In this case, the early stopping limit was set to 20 epochs without any change to performance, with an overall upper limit of 4000 epochs. In order to assess the validation performance, 10% of the created training dataset were used as validation set. After completion of the training, the U-Net’s binarization performance was assessed on the corresponding test dataset.

### C. Measuring Binarization Performance

The pseudo F-Measure ( $F_{ps}$ ) [26], one of the main performance metrics used in DIBCO and H-DIBCO, was used to evaluate the performance of the U-Nets.  $F_{ps}$  is an attempt to address issues with other common evaluation metrics such as PSNR and F-Measure. The main issue with such metrics is that they do not make any difference between errors that impact readability and errors that do not. This is what  $F_{ps}$  attempts to resolve. Much like how F-Measure is based on the precision and recall metrics,  $F_{ps}$  is based on the pseudo precision ( $P_{ps}$ ) and pseudo recall ( $R_{ps}$ ) metrics. These two metrics ensure that errors that impact readability are penalised more than errors that do not by taking into account the location of the mislabelled pixels.  $R_{ps}$  penalises missed pixels in the foreground more if it is close to the middle of a line in the ground truth, while  $P_{ps}$  penalises misclassified background pixels more if they are located close to a line in the ground truth.

## VI. RESULTS AND ANALYSIS

In this section, we evaluate our proposed sample selection method with respect to the reduction in training set size, as well as with respect to the binarization performance achieved by a U-Net trained on the selected dataset. For this evaluation, we compare the performance of a U-Net trained on the whole dataset (*Whole*) with a U-Net trained on a subset selected by our approach using the Euclidean distance for building the RNG (*Euclidean*) and a U-Net trained on a subset selected by our approach using SSIM (*SSIM*).

As shown in Table I and Figure 2, the sample selection approach using SSIM clearly results in the highest reduction

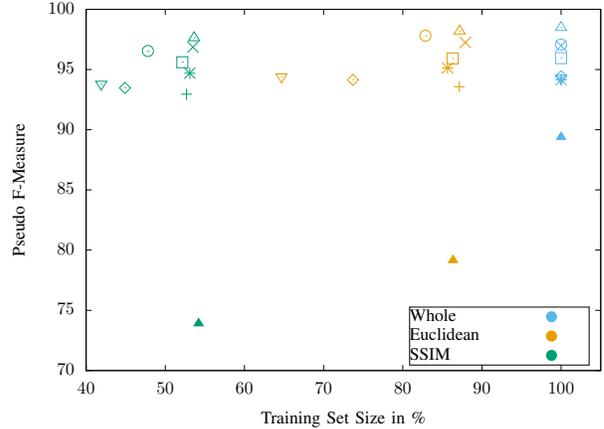


Figure 2. Median binarization performance in  $F_{ps}$  for each test dataset with respect to the training set size in % for the whole training dataset and the selected subsets; symbols indicate the respective test dataset in Table I and colors indicate the respective selection approach.

in training set size with an average size of 50.43%, compared to an average size of 82.42% for the Euclidean distance based approach. This higher reduction in training set size is caused presumably by the fact that the patch labels used to identify the bridge vectors have been generated using SSIM based clustering. Due to the higher alignment between the clusters and the edges in the RNG, less bridge vectors are identified, and thus less training patches are chosen.

While choosing less training samples reduces the human labelling effort, we have to ensure that this reduction in training set size does not have a negative impact on the binarization performance. Therefore, we measure the binarization performance of the U-Nets trained on the respective full or selected training datasets for each of the nine test sets. Figure 3 shows the binarization performance in  $F_{ps}$  for all images from these nine datasets. This figure shows no apparent difference between the selection methods. However when looking at the median performances shown in Figure 2, one can see that *SSIM* tends to produce a lower binarization performance than *Euclidean* and *Whole*. This is particularly clear for the case when the H-DIBCO 2018 dataset [25] is used as test set. Figure 4 illustrates this difference in binarization performance based on one image from this dataset, which shows the largest difference between the binarized images in non-text regions.

This observed difference in binarization performance is confirmed by the performed Friedman test, which indicates that there is a statistically significant difference at the  $p < 0.05$  level between the methods. An afterwards performed pairwise Wilcoxon signed rank test with Holm correction shows that there is a statistically significant difference at the  $p < 0.05$  level between *SSIM* and *Euclidean*, and between *SSIM* and *Whole*. However, there is not statistically significant difference between *Euclidean* and *Whole*. This

Table I  
TRAINING SET SIZES FOR EACH OF THE NINE TEST DATASETS

Dataset:		d09 (+)	d10 (×)	d11 (*)	d12 (□)	d13 (○)	d14 (△)	d16 (▽)	d17 (◇)	d18 (▲)
Whole	# training patches	9182	9135	8914	8351	7697	8958	8766	7449	8668
Euclidean	# training patches	8001	8030	7636	7206	6378	7810	5670	5490	7485
	% of whole	87.14%	87.90%	85.66%	86.29%	82.86%	87.18%	64.68%	73.70%	86.35%
SSIM	# training patches	4836	4885	4732	4355	3679	4805	3674	3345	4698
	% of whole	52.67%	53.48%	53.09%	52.15%	47.80%	53.64%	41.91%	44.91%	54.20%

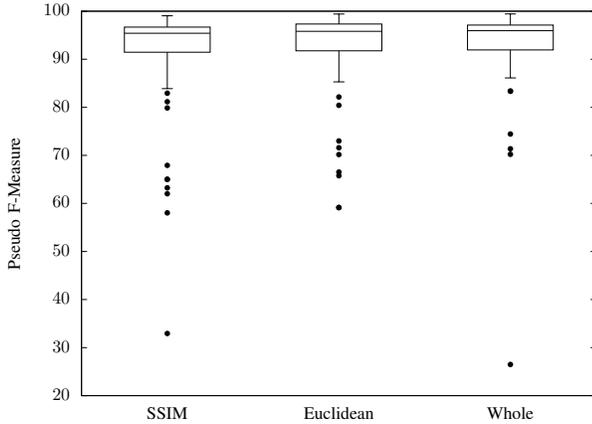


Figure 3. Binarization performance in  $F_{ps}$  over all test images for the whole training dataset and the selected subsets. Boxes show the first, second and third quartile; Whiskers indicate the lowest and highest value within the 1.5 interquartile range; Dots mark outliers outside this range.

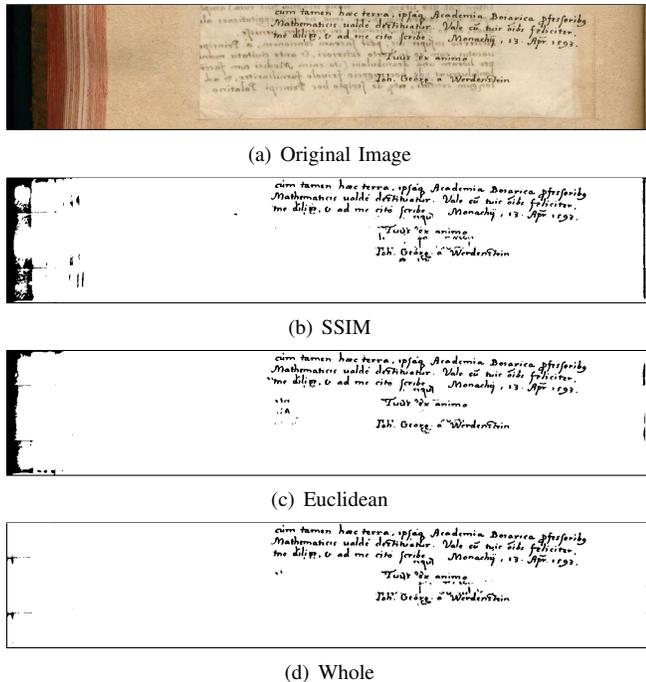


Figure 4. Qualitative comparison between the selected subsets and the whole training dataset for one image of the d18 (▲) dataset [25].

indicates that the sample selection using SSIM consistently reduces the binarization performance in favour for smaller training set sizes. On the other hand, the sample selection approach based on the Euclidean distance does not lead to a significantly lower binarization performance by selecting large parts of the training set.

## VII. CONCLUSIONS

In this paper, we have proposed a sample selection method for document image binarization, which does not depend on available ground truth data to select training samples. This makes it possible to apply the proposed method to identify which patches of a given dataset should be labeled by a human, in order to achieve reasonable binarization performance. We have shown that our proposed approach reduces the size of the used training set by 49.57%, if the structural similarity (SSIM) index is used to construct the relative neighbourhood graph, which is the base for our training sample selection. Furthermore, we have shown that the training set size is reduced by only 17.58% if SSIM is replaced by the commonly used Euclidean distance. While SSIM leads to a greater reduction in training set size, we find that it leads to a statistically significant decrease in binarization performance compared to the whole dataset, as well as the dataset selected using the Euclidean distance. However, since this is a decrease of only 0.55% in median binarization performance over all test sets, we would argue that it is an acceptable trade-off, considering the saved labelling effort.

While the use of SSIM in our sample selection method leads to a reasonable reduction in training set size without resulting in a large decrease in binarization performance, it would be interesting to see if the use of other similarity measures could improve the obtained results. Therefore, we are planning to extend our basic method to use other measures, which may better capture the image characteristics relevant for binarization than SSIM.

## ACKNOWLEDGMENT

This work is part of the research project "Scalable resource-efficient systems for big data analytics" funded by the Knowledge Foundation (grant: 20140032) in Sweden.

## REFERENCES

[1] P. E. Hart, "The Condensed Nearest Neighbor Rule," *IEEE Trans. Inf. Theory*, vol. 14, no. 3, pp. 515–516, 1968.

- [2] F. Rayar, M. Goto, and S. Uchida, "CNN Training with Graph-based Sample Preselection: Application to Handwritten Character Recognition," in *Proc. 13th IAPR Int. Workshop Document Anal. Syst.* Vienna, Austria: IEEE, Jun. 2018, pp. 19–24.
- [3] Z. Zhang, Z. Zhou, and D. Shen, "Sample Selection Method in Supervised Learning Based on Adaptive Estimated Threshold," in *Proc. 12th Int. Conf. Machine Learning and Cybern.*, vol. 4. Tianjin, China: IEEE, 2013, pp. 1861–1864.
- [4] B. K. Bhattacharya, R. S. Poulsen, and G. T. Toussaint, "Application of proximity graphs to editing nearest neighbor decision rules," in *Int. Symp. Inf. Theory.* Santa Monica, USA: IEEE, Feb. 1981.
- [5] M. Goto, R. Ishida, and S. Uchida, "Preselection of Support Vector Candidates by Relative Neighborhood Graph for Large-scale Character Recognition," in *Proc. 13th Int. Conf. Document Anal. and Recognit.* Tunis, Tunisia: IEEE, Aug. 2015, pp. 306–310.
- [6] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proc. 18th Int. Conf. Medical Image Comput. and Computer-Assisted Intervention*, vol. 9351. Munich, Germany: Springer, 2015, pp. 234–241.
- [7] I. Pratikakis, K. Zagoris, G. Barlas, and B. Gatos, "ICDAR2017 Competition on Document Image Binarization (DIBCO 2017)," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. Kyoto: IEEE, Nov. 2017, pp. 1395–1403, 00007. [Online]. Available: <http://ieeexplore.ieee.org/document/8270159/>
- [8] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Trans. on Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004. [Online]. Available: <http://ieeexplore.ieee.org/document/1284395/>
- [9] B. J. Frey and D. Dueck, "Mixture Modeling by Affinity Propagation," in *Advances in Neural Inf. Process. Syst.* Vancouver, Canada: NIPS, Dec. 2005, pp. 379–386.
- [10] —, "Clustering by Passing Messages Between Data Points," *Science*, vol. 315, no. 5814, pp. 972–976, Feb. 2007. [Online]. Available: <http://www.sciencemag.org/cgi/doi/10.1126/science.1136800>
- [11] D. Dueck and B. J. Frey, "Non-metric Affinity Propagation for Unsupervised Image Categorization," in *Proc. 11th Int. Conf. Comput. Vision.* Rio de Janeiro, Brazil: IEEE, Oct. 2007, pp. 1–8.
- [12] G. T. Toussaint, "The Relative Neighbourhood Graph of a Finite Planar Set," *Pattern Recognit.*, vol. 12, no. 4, pp. 261–268, Jan. 1980. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0031320380900667>
- [13] L. Wang, Y. Zhang, and J. Feng, "On the Euclidean Distance of Images," *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 27, no. 8, pp. 1334–1339, Aug. 2005.
- [14] F. Chollet and others, "Keras," 2015, <https://keras.io>. [Online]. Available: <https://keras.io>
- [15] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine Learning in Python," *J. Machine Learning Res.*, vol. 12, pp. 2825–2830, Oct. 2011. [Online]. Available: <http://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html>
- [16] A. A. Hagberg, D. A. Schult, and P. J. Swart, "Exploring Network Structure, Dynamics, and Function Using NetworkX," in *Proc. 7th Python in Sci. Conf.*, G. Varoquaux, T. Vaught, and J. Millman, Eds., Pasadena, CA USA, 2008, pp. 11 – 15.
- [17] B. Gatos, K. Ntirogiannis, and I. Pratikakis, "ICDAR 2009 Document Image Binarization Contest (DIBCO 2009)," in *Proc. 10th Int. Conf. Document Anal. and Recognit.* Barcelona, Spain: IEEE, Jul. 2009, pp. 1375–1382.
- [18] I. Pratikakis, B. Gatos, and K. Ntirogiannis, "ICDAR 2011 Document Image Binarization Contest (DIBCO 2011)," in *Proc. 11th Int. Conf. Document Anal. and Recognit.* Beijing, China: IEEE, Sep. 2011, pp. 1506–1510.
- [19] —, "ICDAR 2013 Document Image Binarization Contest (DIBCO 2013)," in *Proc. 12th Int. Conf. Document Anal. and Recognit.* Washington, DC, USA: IEEE, Aug. 2013, pp. 1471–1476.
- [20] I. Pratikakis, K. Zagoris, G. Barlas, and B. Gatos, "ICDAR2017 Competition on Document Image Binarization (DIBCO 2017)," in *Proc. 14th IAPR Int. Conf. Document Anal. and Recognit.*, vol. 1. Kyoto, Japan: IEEE, Jan. 2018, pp. 1395–1403.
- [21] I. Pratikakis, B. Gatos, and K. Ntirogiannis, "H-DIBCO 2010 - Handwritten Document Image Binarization Competition," in *Proc. 12th Int. Conf. Frontiers in Handwriting Recognit.* Kolkata, India: IEEE, Nov. 2010, pp. 727–732.
- [22] —, "ICFHR 2012 Competition on Handwritten Document Image Binarization (H-DIBCO 2012)," in *Proc. 13th Int. Conf. Frontiers in Handwriting Recognit.* Bari, Italy: IEEE, Sep. 2012, pp. 817–822.
- [23] K. Ntirogiannis, B. Gatos, and I. Pratikakis, "ICFHR2014 Competition on Handwritten Document Image Binarization (H-DIBCO 2014)," in *Proc. 14th Int. Conf. Frontiers in Handwriting Recognit.* Heraklion, Greece: IEEE, Sep. 2014, pp. 809–813.
- [24] I. Pratikakis, K. Zagoris, G. Barlas, and B. Gatos, "ICFHR2016 Handwritten Document Image Binarization Contest (H-DIBCO 2016)," in *Proc. 15th Int. Conf. Frontiers in Handwriting Recognit.* Shenzhen, China: IEEE, Oct. 2016, pp. 619–623. [Online]. Available: [doi.ieeecomputersociety.org/10.1109/ICFHR.2016.0118](https://doi.org/10.1109/ICFHR.2016.0118)
- [25] I. Pratikakis, K. Zagori, P. Kaddas, and B. Gatos, "ICFHR 2018 Competition on Handwritten Document Image Binarization (H-DIBCO 2018)," in *Proc. 16th Int. Conf. Frontiers in Handwriting Recognit.* Niagara Falls, NY, USA: IEEE, Aug. 2018, pp. 489–493.
- [26] K. Ntirogiannis, B. Gatos, and I. Pratikakis, "Performance Evaluation Methodology for Historical Document Image Binarization," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 595–609, Feb. 2013.