

Harnessing Neural Networks for Enhancing Image Binarization Through Threshold Combination

Giorgiana Violeta VLĂSCEANU¹,
Nicolae TARBĂ²

¹Teaching assistant, PhD student, Eng.,
University Politehnica of Bucharest, 060042
Bucharest, Romania,

giorgiana.vlasceanu@cs.pub.ro

² PhD student, Eng., University Politehnica
of Bucharest, 060042 Bucharest, Romania,
nicolae.tarba@upb.ro

Abstract: *Threshold-based methods are prevalent across numerous domains, with specific relevance to image binarization, which traditionally employs global and local threshold algorithms. This paper presents a novel approach to image binarization, where the capacity of neural networks is utilized not just for determining optimal thresholds, but also for combining multiple global thresholds sourced from existing binarization techniques. The primary objective of our method is to develop a robust binarization strategy capable of managing a wide array of image conditions. By integrating the strengths of various thresholding techniques, our approach aims to establish a significant connection between traditional thresholding methods and those underpinned by deep learning.*

Keywords: *threshold, image binarization, neural network, global thresholding, multi-thresholding combination.*

How to cite: Vlăsceanu, G. V., Tarbă, N. (2023). Harnessing Neural Networks for Enhancing Image Binarization Through Threshold Combination. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, 14(2), 59-75.
<https://doi.org/10.18662/brain/14.2/444>

1. Introduction

Image binarization, the conversion of a grayscale or color image into a binary image, has been a cornerstone in the field of computer vision for decades. It finds widespread applications in optical character recognition (OCR), document analysis, image segmentation, and beyond. Traditionally, this process involves the application of a threshold value to distinguish between the foreground and background of the image.

Various thresholding algorithms, ranging from simple global thresholding to adaptive local techniques, have been proposed in the literature. Examples include Otsu's method (1979) in the global approach and for local Sauvola's method (Sauvola & Pietikäinen, 2000), which have demonstrated effective performance in different application scenarios. However, selecting an appropriate threshold that is versatile across diverse image types and quality remains a challenging task.

In recent years, the advent of deep learning has unlocked a wealth of opportunities in the realm of image binarization. Convolutional neural networks (CNNs) and other deep learning architectures have shown remarkable proficiency in learning complex representations, providing us with the tools to tackle the intricacies of image binarization.

This paper presents a novel approach to image binarization, where we leverage the capabilities of neural networks not just for learning an optimal threshold, but for combining multiple thresholds derived from existing binarization methods. The goal of this approach is to provide a more flexible and resilient binarization strategy that can handle a wider range of image conditions. Incorporating the strength of multiple thresholding techniques, our proposed method aims to bridge the gap between traditional and deep learning-based binarization methods. By training the network to make judicious use of different thresholds, we aspire to bring a new level of flexibility and adaptability to the process of image binarization.

In the ensuing sections, we shall delve into the details of our approach, beginning with a brief review of existing binarization techniques and neural network models relevant to our work. We shall then describe the architecture of our neural network model and the method of combining thresholds. Finally, we will evaluate the performance of our proposed approach through rigorous experimental validations, demonstrating its efficacy and potential.

2. Related work

The process of image binarization has been extensively studied, with a plethora of methods proposed to perform this task efficiently. Traditional methods primarily focus on selecting an appropriate threshold for binarization, while contemporary approaches employ more advanced deep-learning models to perform this task.

Within the realm of classical methods for image binarization, a significant area of interest is the domain of global thresholding algorithms. The underlying principle of these methods is the computation of a single, universal threshold that is applied uniformly across the entire image. This approach operates on the assumption that the image has a distinct division between the foreground and the background, which can be differentiated based on their intensity levels.

Global thresholding algorithms have their roots in the early developments of image processing and continue to hold relevance due to their simplicity and computational efficiency. They are particularly effective when the image exhibits a bimodal histogram, i.e., distinct peaks corresponding to the foreground and background.

Among these global thresholding methods, several have gained remarkable recognition for their effectiveness and wide applicability. This includes the following methods: Otsu (1979), Kittler & Illingworth (1985), Lloyd (1985), Sung et al. (2014), Ridler & Calvard (1978), Huang & Wang (1995), Ramesh et al. (1995), two variances of Li & Lee (1993), Brink & Pendock (1996), Kapur et al. (1985), Sahoo et al. (1997), Shanbha (1994), Yen et al. (1995), Tsai (1985).

While these methods offer significant advantages, they are not devoid of limitations, particularly when dealing with images that have uneven illumination or lack a distinct separation between the foreground and the background. However, their contribution to the field of image binarization cannot be understated. In section 3.1 we delve deeper into the exploration of these global thresholding algorithms in order to showcase the input values for the proposed solution.

Modern strategies for image binarization encompass a broad understanding of the inherent complexities of the task. These contemporary methods leverage advanced technologies, ranging from deep learning to traditional machine learning algorithms. This rich toolset provides versatile solutions that cater to the unique demands of image binarization.

We investigate a variety of systems specifically designed for image binarization in our study. It's important to note that the application of these

methods extends beyond document images. They are also widely utilized in specialized domains such as medical imaging and biological studies. For example, in (Kodieswari, 2022) the authors proposed a solution that seeks to address current challenges in lung cancer detection. It begins by feeding human lung CT scans into a preprocessing stage, which is then followed by binarization to yield a binary image for cancer detection. The image is subsequently segmented, and each segment undergoes thorough analysis through feature extraction. They used a Convolutional Neural Network (CNN) to classify the identified tumor cells into malignant or benign categories with remarkable results - an accuracy rate of 95.

In Xu et al. (2019) the solution proposed utilizes a deep learning model called U-Net for detecting biomarkers in medical images. It is known that U-Net has a tendency to undersegment the input without proper intensity thresholding. To mitigate this, the paper suggests combining U-Net with optimal Otsu thresholding. This approach demonstrated promising results.

Another interesting work comes from the field of drone imagery (Zhu et al., 2021). The paper introduces a deep learning method for the automated inspection of infrastructure defects using drones (Zhu et al., 2021). It employs hierarchical convolutional neural networks with feature preservation and iterative intercontrast thresholding for image binarization. The process involves integrating outputs from previous and current convolutional blocks to reduce information loss during down-sampling. A Contrast-Based Autotuned Thresholding method is then used to extract and cluster features. This technique proves effective for detecting surface cracks on roads, bridges, and pavements.

Moving to document image binarization, alternative strategies employ classification-based approaches. Hamza et al. (2005) and Kefali et al. (2014) implemented a Multi-Layer Perceptron classifier (MLP), with the former using pixel labels from clustering and the latter classifying pixels based on the surrounding intensity values and overall image statistics. Afzal et al.'s strategy involved a 2D Long Short Term Memory (LSTM) network (Afzal et al., 2015), which notably reduced Optical Character Recognition (OCR) errors in comparison to Sauvola's method (Sauvola & Pietikäinen, 2000).

Approaches utilizing Convolutional Neural Networks (CNNs) have also been investigated. Pastor-Pellicer et al., for instance, used a CNN to classify pixels considering their surrounding intensity values (Pastor-Pellicer et al., 2015). Meanwhile, Wu et al. (2016) employed an Extremely Randomized Trees classifier that was trained utilizing an array of statistical

and heuristic properties retrieved from the pixel. For the broader challenge of semantic segmentation in natural images, Long et al. proposed a Fully Convolutional Networks (FCNs) (Li et al., 2020). Zheng and Chen have tried to improve prediction localization and consistency by combining FCNs with Conditional Random Fields (CRFs) (Chen et al., 2018; Zheng et al., 2015). A breakthrough in the domain was brought by Tensmeyer et al with their approach to “binarization as a pixel classification” manner and a Fully Convolutional Network (FCN) architecture applied (Tensmeyer & Martinez, 2017). The FCN is taught to optimize an ongoing iteration of the Pseudo F-Measure metric, with a group of FCNs outperforming the winners of four of seven DIBCO competitions.

Kang et al proposed a scenario where the aforementioned U-net model is integrated into a cascading construction called Cascading Modular U-Nets (CMU-Nets) (Kang et al., 2021). CMU-Nets comprise pre-trained modular components, offering a solution to the challenge posed by a limited quantity of training images. U-net model was used also in (He & Schomaker, 2019) to understand the deteriorations present in document images.

Calvo-Zaragoza and Gallego, introduce a solution that involves training a Selectional Auto-Encoder (SAE) (2019). The SAE can learn from start-to-finish conversion for image binarization using this method. Westphal et al use RNN to employ Grid LSTM to manage multidimensional input (2018). Another example of CNNs uses (CNNs) with multichannel images as input, employing wavelet analysis (Akbari et al., 2020).

3. Approach of the proposed system

This work introduces an innovative strategy for image binarization, where the power of neural networks is harnessed in a unique way. Instead of merely employing neural networks for discerning an optimal threshold value, we utilize them to amalgamate multiple threshold values which are derived from a range of existing image binarization techniques. This not only introduces a level of adaptability in our approach but also builds upon the collective strengths of established methods, hence offering a more comprehensive solution to the image binarization problem. This underscores the versatility of neural networks in handling complex tasks and presents an inventive angle to the image binarization process.

3.1. The Employed Algorithms

Building upon our prior discussion, the neural network operates on a set of 15 distinct threshold values. These values are meticulously derived from a diverse collection of established algorithms that feature prominently

in academic literature. In the following section, we will delve into a comprehensive introduction to these individual thresholding algorithms, shedding light on their function and significance in this context.

The **Otsu** method (1979), a renowned global thresholding technique, is used to transform grayscale images into binary images by separating the foreground from the background based on intraclass variance minimization. The method calculates a threshold minimizing the weighted within-class variance in an image's grayscale histogram. The technique presumes the image contains two-pixel classes, foreground, and background, with a bimodal histogram, and it computes the optimal threshold separating these classes to minimize their combined spread or maximize their inter-class variance. The Otsu method's main advantage lies in its simplicity and efficiency, especially for images with a clear bimodal gray-level histogram. However, its performance may wane in complex situations where the two-pixel classes assumption doesn't apply, or when the histogram isn't bimodal. Despite these limitations, Otsu's method remains a critical tool in image processing, specifically in image binarization.

The **Kittler & Illingworth** thresholding method (1985), also known as Minimum Error Thresholding, is a common approach in image binarization. It establishes the threshold by framing the task as a statistical classification problem, minimizing the classification error between foreground and background pixel classes. The method employs an iterative process that uses a cost function based on class membership probabilities and misclassification cost. The cost function measures the number of misclassified pixels from each class, adjusted by the cost of misclassification. The algorithm cycles through potential threshold values and selects the one that minimizes the cost function. Distinct from Otsu's method which seeks to maximize between-class variance, Kittler's approach minimizes a cost function reflecting direct classification error. While it offers a robust thresholding strategy, especially useful for unimodal histograms or significantly noisy images, its efficacy can depend on the assumptions about pixel classes' statistical distributions and misclassification costs.

Lloyd's method (1985) minimizes the mean squared quantization error to find the optimal threshold. Essentially, it works to lessen the variance between the original grayscale image and the final binary outcome. It operates iteratively, continuously adjusting the threshold value to minimize the mean squared error between the original image's gray levels and the assigned binary class values. Each iteration assigns each pixel to the nearest threshold and recalculates the threshold as the mean gray level of assigned pixels. The process iterates until the thresholds stabilize. Despite its

computational intensity due to the iterative nature of the method, Lloyd's approach can provide high-quality results, especially for images with intricate gray-level distributions.

Sung et al. (2014) introduces a threshold selection criterion based on the within-class standard deviation. The optimal threshold is identified by minimizing this standard deviation. The experimental findings support the superiority of this method over existing algorithms in terms of performance. Put differently, this method reduces the bias of the optimal threshold.

Ridler's & Calvard thresholding (1978), also known as the Iterative Self-Organizing Data Analysis Technique Algorithm (ISODATA), is a robust method for image binarization. The technique is based on an iterative process that successively refines an initially guessed threshold. The method starts with a preliminary threshold, often the midpoint of the grayscale range, and then calculates two means: one for all the pixels above the threshold, and one for those below. The next threshold is then determined by taking an average of these two means. This iterative process is performed until the threshold value stabilizes. The strength of Ridler's method lies in its simplicity and efficiency, making it suitable for a variety of applications in image processing and computer vision.

Huang's & Wang thresholding (1995) applies a fuzzy partitioning of the histogram, utilizing the concept of fuzzy entropy. It measures the "fuzziness" of an image to determine the threshold value. This method is found to be particularly effective in images with blurred boundaries, as the fuzziness concept allows for some level of uncertainty and ambiguity, which is often the case with these images is an entropy-based method for calculating the threshold by maximizing the entropy within the image's foreground and background.

Ramesh et al. thresholding (1995) is another entropy-based technique, but it extends upon Huang's method by incorporating a correction factor into the entropy calculation to account for variations in the histogram. This correction factor improves the method's robustness, particularly in the presence of noise or when the histogram has irregular peaks.

Li's & Lee thresholding method (1993), also known as the Cross-Entropy Thresholding, minimizes the cross-entropy between the original and thresholded images to optimize the threshold. This method offers a robust alternative to entropy-based thresholding methods, capable of handling a broader range of image conditions and histogram distributions. In the presented solution, we have two variations of this method - Li1 and Li2. Li1 uses an iterative approach to refine the threshold value and keeps track

of foreground and background data for each potential threshold, which is computationally intensive but potentially more accurate. It also applies an offset for more accurate averages and log values in both the threshold refinement and confidence calculation. In contrast, Li2 calculates the threshold in a single pass by minimizing an optimality measure for each potential threshold. It calculates foreground and background data on-the-fly without storing it, and applies the offset only in the optimality measure calculation. This approach is potentially less accurate but more efficient. Both methods calculate a final confidence value, albeit differently. The choice between the two would depend on specific application requirements.

Brink's & Pendock thresholding algorithm (1996) focuses on histogram shape, rather than pixel intensity values. It applies a gradient descent strategy on the histogram curve, starting from the peak and moving towards the lower-valued tail until a "valley" is found. This valley position represents the threshold. Brink's method is particularly effective for bimodal histograms where the two modes have significantly different peaks.

Kapur's et al. thresholding (1985), also known as Maximum Entropy Thresholding, differs from Huang's and Ramesh's methods by increasing the sum of both the foreground and background entropies. This method assumes that a well-separated image will yield maximum information, thereby achieving optimal binarization. Kapur's approach has been widely adopted in applications dealing with multimodal or complex histograms.

Sahoo's et al. thresholding technique (1997) is an entropy-based method that seeks to maximize the entropy of the histogram. Unlike other entropy-based methods, Sahoo's technique calculates the entropy of the gray level probabilities before and after the threshold, then adds these two entropy values together. The optimal threshold is the one that maximizes this sum. This method is particularly useful in scenarios where the image histogram is not clearly bimodal, providing a robust solution for images with complex intensity distributions.

Shanbhag's thresholding (1994) is a unique approach based on the principle of moments preservation. The algorithm estimates the threshold by ensuring that the moments before and after thresholding remain the same. This method has been shown to be effective in various contexts, including images with complicated intensity distributions and noisy conditions.

Yen's et al. thresholding method (1995), a derivative of the Maximum Entropy Thresholding, in which the threshold is calculated by maximizing the normal difference between the two sub-histograms produced by the thresholding process. By maximizing this normal difference, Yen's method ensures that the threshold effectively separates the

foreground and background of the image. It is recognized for its strong performance in handling a range of image conditions.

Tsai's thresholding method (1985) is a versatile technique that operates based on the moment-preserving principle. Tsai's approach calculates the threshold by ensuring the preservation of moments in the thresholded image. This method has shown notable effectiveness in scenarios where the intensity distribution is complex or where noise is present, providing a practical tool for a variety of image binarization tasks.

3.2. Datasets involved in the process

We calculated the thresholds by employing a diverse selection of reputable datasets known for their use in image binarization research. The datasets incorporated in this process include:

1. **DIBCO (Document Image Binarization COmpetition):** DIBCO (n.d.) is a popular dataset used in binarization competitions, containing a wide variety of document images with different types of noise and degradation. We used images from competitions between 2009 and 2019.
2. **H-DIBCO (Hellenic DIBCO)** (n.d.): This dataset is a variant of DIBCO, also used in image binarization competitions. It consists of handwritten documents with different types of noise.
3. **NoisyOffice** (n.d.): Benchmark dataset that is used for evaluating document image binarization methods. The dataset is comprised of document images with different types of artificial noise added to them, such as salt-and-pepper noise, Gaussian noise, and speckle noise, among others.
4. **The PHIBD** (Printed Historical Indian Books Dataset) (n.d.) is an extensive collection of images from historical Indian printed books. It offers a comprehensive platform for evaluating image binarization methods, given its diverse content including different languages, scripts, and degrees of degradation and noise.
5. **BICKLEY DIARY** (Su et al., 2013): This dataset contains images of the Bickley Diary, a handwritten historical document. It is often used to evaluate the performance of binarization techniques on historical documents.
6. **Palm Leaf** (Kingma & Ba, 2014) includes images of ancient palm leaf manuscripts. This dataset, characterized by complex backgrounds, fragmented character strokes, and varied leaf

textures and colors, provides a challenging platform to evaluate binarization methods against variable and difficult conditions.

7. **Nabucco** is a dataset that includes 6500 letters and postcards. The images contain different alterations.

Our study utilized a collected dataset comprising 1,195 images. These images were pre-processed and adjusted with gamma correction (the value for gamma is increased with 0.1 step in interval 0.5 – 2) and obtained 19,120 images. The dataset obtained is employed for both the training and testing phases. For each image, a set of 15 thresholding values was computed utilizing the suite of algorithms previously delineated. Furthermore, a proprietary optimal threshold value was computed using a solution designed by our research team, serving as the ground truth for our system.

To ensure rigorous testing and evaluation, we reserved 30% of the entire dataset as a test set. This partitioning strategy facilitates the validation of our model's learning efficacy and its capacity to generalize beyond the training data. It also allows us to quantitatively assess the performance of our system on unseen data, thereby providing a comprehensive view of the system's reliability and robustness.

3.3. The deployed Neural Network

The architecture under review constitutes a feed-forward artificial neural network, implemented via a Sequential model using Keras¹ and TensorFlow². This model enables the linear stacking of layers, where each layer has a single input tensor and a single output tensor.

The initial layer in this model is a fully connected or Dense layer with 512 neurons. It employs the Rectified Linear Unit (ReLU) activation function, widely acknowledged for its capacity to counter the vanishing gradient problem prevalent in deep neural networks. The layer's input dimension is determined by the 15 features in the dataset. The model further incorporates two subsequent Dense layers, each containing 512 neurons and using the ReLU activation function. These layers autonomously infer their input dimensions from their predecessor, obviating the need for explicit specifications. The final layer of the network architecture is a Dense layer equipped with a single unit and a sigmoid activation function. The choice of the sigmoid function is particularly suitable for our case since our features are normalized, and we aim for the network's output to be a continuous value within the range of 0 to 1.

¹ <https://keras.io>

² <https://www.tensorflow.org>

The training regimen for the neural network is established with a specification of 50 epochs with a validation split of 0.2 and a batch size accommodating 500 sets of 15 features each – equivalent to 500 images. These parameters, while fixed for the current study, offer scope for hyperparameter tuning to potentially further enhance the model's performance. The training is conducted using the Mean Squared Error (MSE) as the loss function, which quantifies the deviation between the predicted and actual values. Lower MSE values signify better model performance, as they indicate that the model's predictions closely align with the actual values. Therefore, our primary objective during training is to minimize this error metric, thereby ensuring accurate and reliable predictions from our model. The neural network model's compilation process employs the Adam optimizer, a widely accepted optimization algorithm (Kingma & Ba, 2014).

For each image in the dataset, we construct a lookup table that details the F-measure for various thresholding intervals based on the image histogram. This method provides a straightforward mechanism for evaluating the F-measure associated with any predicted threshold value. This approach allows us to validate and assess the efficacy of the predicted thresholds directly and efficiently.

4. Evaluation

We applied the trained model to our test dataset, and a selection of 200 predicted values is illustrated in Figure 1. This visualization clearly reveals the alignment between our model's predicted values and the target values, demonstrating the model's ability to accurately estimate thresholds. The adherence of the predicted outcomes to the target values underscores the efficacy of our proposed neural network architecture in image binarization tasks.

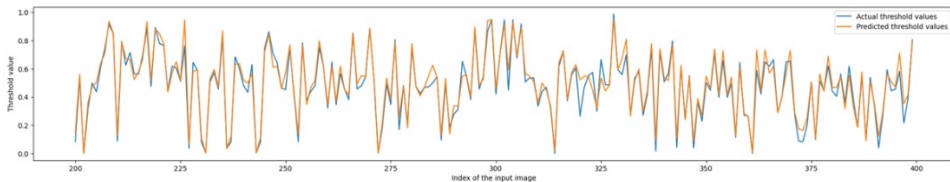


Figure 1. Thresholding values for 200 images in the dataset
Source: Author's own conception

The effectiveness of our neural network is quantified by calculating the mean squared error (MSE), a popular metric used to measure the average

of the squares of the errors. In this case, the MSE value yielded by our model is 0.03. This relatively low score indicates the model's effectiveness, demonstrating its capability to predict thresholds with a high degree of accuracy. The network loss is presented in Figure 2.

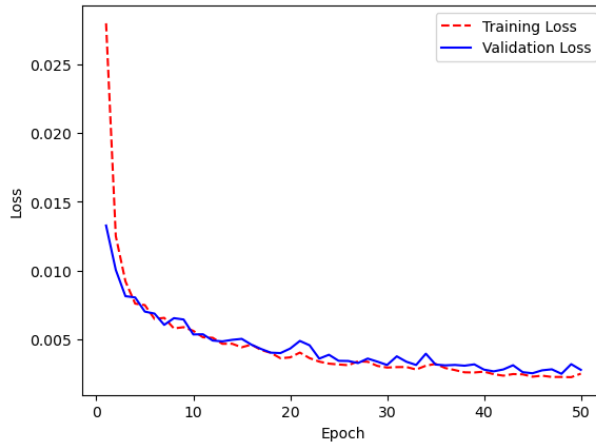


Figure 2. Loss history of the model
Source: Author's own conception

The performance of the proposed regression neural network has been evaluated using several key metrics, with encouraging results observed across all measures, as presented in Table 1. The mean squared error (MSE), a fundamental measure of regression model performance, came out to be as low as 0.003. This suggests that our model predictions deviate very little from the actual values, highlighting its accuracy. Complementing the MSE, the root mean square error (RMSE) was calculated to be 0.054. The RMSE is particularly useful as it gives the deviations in the units of the target variable, thereby providing a more interpretable measure of error magnitude. The low RMSE value indicates a strong predictive performance of the network, with a low spread of the residuals.

Furthermore, the model achieved an impressive R2 score of 0.94, indicating that our model explains 94% of the variance in the dependent variable that is predictable from the independent variables. This is a testament to the model's excellent goodness of fit.

Lastly, the mean absolute error (MAE) was found to be 0.036. MAE provides a direct interpretation of how far off the predictions are from the actual values on average. Our model's low MAE suggests that, on average, the predicted threshold deviates only slightly from the actual value, reinforcing the strong performance of our proposed approach.

Altogether, these metrics provide substantial evidence of the proposed neural network's robust performance in predicting the image binarization thresholds.

	MSE	RMSE	R2	MAE
Proposed Model	0.003	0.054	0.948	0.036

Table 1. Evaluation scores for the neural network
Source: Author's own conception

For the thresholding evaluation, we applied our model across the entire dataset to generate predicted threshold values. Leveraging the lookup table, we averaged the F-Measure of these predicted thresholds, achieving a result of 77.34. Considering we utilized exclusively global thresholding methods, this result is quite satisfactory.

Method	Ideal	Otsu	Kittler	Lloyd	Sung	Ridler	Huang
F-Measure	81.75	67.11	64.56	63.65	61.68	64.88	53.06

Ramesh	Li1	Li2	Brink	Kapur	Sahoo	Shabang	Yen	Tsai
53.67	66.51	68.59	62.33	60.13	59.91	50.22	58.46	60.24

Table 2. Scores for individual methods in the used dataset
Source: Author's own conception

As a point of comparison, the maximal value for optimal global thresholding is approximately 81.75% in the dataset we used, as presented in Table 2. The results predicted by the proposed solution surpass the individual method. Thus, our model's performance stands up favorably within this context.

5. Conclusions

This study has successfully demonstrated the potential of neural networks in enhancing the robustness and adaptability of image binarization, a fundamental process in computer vision. By innovatively integrating multiple global thresholding techniques into the learning process, we've managed to navigate the complexity and variability of image conditions, underlining the pivotal role of deep learning methodologies. Our model, trained on a range of diverse datasets and evaluated with robust metrics, has achieved satisfactory results. Going forward, we aim to continue refining our approach and exploring more sophisticated architectures and algorithms to further elevate the process of image binarization.

Acknowledgment

The results presented in this article has been funded by the Ministry of Investments and European Projects through the Human Capital Sectoral Operational Program 2014-2020, Contract no. 62461/03.06.2022, SMIS code 153735.

References

- Afzal, M. Z., Pastor-Pellicer, J., Shafait, F., Breuel, T. M., Dengel, A., & Liwicki, M. (2015, August 22). Document image binarization using LSTM: A sequence learning approach. *Proceedings of the 3rd International Workshop on Historical Document Imaging and Processing*. Tunisia.
<https://doi.org/10.1145/2809544.280956>
- Akbari, Y., Al-Maadeed, S., & Adam, K. (2020). Binarization of Degraded Document Images Using Convolutional Neural Networks and Wavelet-Based Multichannel Images. *IEEE Access*, 8, 153517–153534.
<https://doi.org/10.1109/ACCESS.2020.3017783>
- Brink, A. D., & Pendock, N. E. (1996). Minimum cross entropy intThreshold selection. *Pattern Recognition*, 29, 179-188.
<https://doi.org/10.1109/SNPD.2007.85>
- Calvo-Zaragoza, J., & Gallego, A-J. (2019). A selectional auto-encoder approach for document image binarization. *Pattern Recognition*, 86, 37–47.
<https://doi.org/10.1016/j.patcog.2018.08.011>
- Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4), 834–848.
<https://doi.org/10.1109/TPAMI.2017.2699184>

- DIBCO Dataset. (n.d.). *About*. DIBCO Dataset.
<https://dib.cin.ufpe.br/#!/resources/dibco>
- Hamza, H., Smigiel, E., & Belaid, E. (2005). Neural based binarization techniques. In *ICDAR 2005* (pp. 317–321). IEEE.
- He, S., & Schomaker, L. (2019). DeepOtsu: Document Enhancement and Binarization using Iterative Deep Learning. *arXiv:1901.06081*.
<https://doi.org/10.48550/arXiv.1901.06081>
- Huang, L.-K., & Wang, M.-J. J. (1995). Image thresholding by minimizing the measures of fuzziness. *Pattern Recognition*, 28(1), 41–51.
[https://doi.org/10.1016/0031-3203\(94\)E0043-K](https://doi.org/10.1016/0031-3203(94)E0043-K)
- Kang, S., Iwana, B. K., & Uchida, S. (2021). Complex image processing with less data—Document image binarization by integrating multiple pre-trained U-Net modules. *Pattern Recognition*, 109, 107577.
<https://doi.org/10.1016/j.patcog.2020.107577>
- Kapur, J. N., Sahoo, P. K., & Wong, A. K. C. (1985). A new method for gray-level picture thresholding using the entropy of the histogram. *Computer Vision, Graphics, and Image Processing*, 29(3), 273–285.
[https://doi.org/10.1016/0734-189X\(85\)90125-2](https://doi.org/10.1016/0734-189X(85)90125-2)
- Kefali, A., Sari, T., & Bahi H. (2014). Foreground-background separation by feedforward neural networks in old manuscripts. *Informatica*, 38(4), 329–338.
<https://www.informatica.si/index.php/informatica/article/view/715/585>
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv:1412.6980*. <https://doi.org/10.48550/arXiv.1412.6980>
- Kittler, J., & Illingworth, J. (1985). On threshold selection using clustering criteria. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-15(5), 652–655. <https://doi.org/10.1109/TSMC.1985.6313443>
- Kodieswari, N, G., Deepa, & Kavitha. (2022, December 10). Analysis and classification of the lung cancer with CNN implementation. *2022 Smart Technologies, Communication and Robotics (STCR)*, India.
<https://doi.org/10.1109/stcr55312.2022.10009558>
- Li, C. H., & Lee, C. K. (1993). Minimum cross entropy thresholding. *Pattern Recognition*, 26(4), 617–625. [https://doi.org/10.1016/0031-3203\(93\)90115-D](https://doi.org/10.1016/0031-3203(93)90115-D)
- Li, F., Long, Z., He, P., Feng, P., Guo, X., Ren, X., & Tang, B. (2020). Fully convolutional pyramidal networks for semantic segmentation. *IEEE Access: Practical Innovations, Open Solutions*, 8, 229132–229140.
<https://doi.org/10.1109/access.2020.3045280>
- Lloyd, D. E. (1985). Automatic target classification using moment invariant of image shapes. *IDN AW126, RAE, Farnborough, Reino Unido*.

- NoisyOffice Dataset. (n.d.). *About. NoisyOffice Dataset.*
<https://archive.ics.uci.edu/ml/datasets/NoisyOffice>
- Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), 62–66.
<https://doi.org/10.1109/TSMC.1979.4310076>
- Pastor-Pellicer, J., España-Boquera, S., Zamora-Martínez, F., Afzal, M. Z., & Castro-Bleda, M. J. (2015). Insights on the use of convolutional neural networks for document image binarization. In *Lecture Notes in Computer Science. Advances in Computational Intelligence* (pp. 115–126).
https://doi.org/10.1007/978-3-319-19222-2_10
- PHIBD 2012 Dataset. (n.d.). *About. PHIBD 2012 Dataset.*
http://www.iapr-tc11.org/mediawiki/index.php/Binarization_of_PHIBD_2012_dataset
- Ramesh, N., Yoo, J. H., & Sethi. I. K. (1995). Thresholding based on histogram approximation. *IEEE Proceeding Vision Image Signal Process*, 142(5), 271–279.
<https://doi.org/10.1049/ip-vis:19952007>
- Ridler, T. W., & Calvard, S. (1978). Picture Thresholding Using an Iterative Selection Method. (1978). *IEEE Transactions on Systems, Man, and Cybernetics*, 8(8), 630–632. <https://doi.org/10.1109/TSMC.1978.4310039>
- Sahoo, P., Wilkins, C., & Yeager, J. (1997). Threshold selection using Renyi's entropy. *Pattern Recognition*, 30(1), 71–84. [https://doi.org/10.1016/s0031-3203\(96\)00065-9](https://doi.org/10.1016/s0031-3203(96)00065-9)
- Sauvola, J., & Pietikäinen, M. (2000). Adaptive document image binarization. *Pattern Recognition*, 33(2), 225–236. [https://doi.org/10.1016/S0031-3203\(99\)00055-2](https://doi.org/10.1016/S0031-3203(99)00055-2)
- Shanhbag, A. G. (1994). Utilization of Information Measure as a Means of Image Thresholding. *Graphical Models and Image Processing*, 56, 414–419.
<https://doi.org/10.1006/cgip.1994.1037>
- Su, B., Lu, S., & Tan, C. L. (2013). Robust Document Image Binarization Technique for Degraded Document Images. *IEEE Transactions on Image Processing*, 22(4), 1408–1417. <https://doi.org/10.1109/TIP.2012.2231089>
- Sung, J.-M., Kim, D.-C., Choi, B.-Y., & Ha, Y.-H. (2014, March 7). Image thresholding using standard deviation. In K. S. Niel & P. R. Bingham (Eds.), *Image Processing: Machine Vision Applications VII*.
<https://doi.org/10.1117/12.2040990>
- Tensmeyer, C., & Martinez, T. (2017, November). Document image binarization with fully convolutional neural networks. *2017 14th LAPR International Conference on Document Analysis and Recognition (ICDAR)*. ICDAR.
- Tsai, W.-H. (1985). Moment-preserving thresholding: A new approach. *Computer Vision, Graphics, and Image Processing*, 29, 377–393.
[https://doi.org/10.1016/0734-189X\(85\)90133-1](https://doi.org/10.1016/0734-189X(85)90133-1)

- Westphal, F., Lavesson, N., & Grahn, H. (2018). Document Image Binarization Using Recurrent Neural Networks. *13th LAPR International Workshop on Document Analysis Systems (DAS)*. <https://doi.org/10.1109/DAS.2018.71>
- Wu, Y., Natarajan, P., Rawls, S., & AbdAlmageed, W. (2016, September). Learning document image binarization from data. *2016 IEEE International Conference on Image Processing (ICIP)*. USA. <https://doi.org/10.1109/icip.2016.7533063>
- Xu, Y., Gao, F., Wu, T., Bennett, K. M., Charlton, J. R., & Sarkar, S. (2019, August). U-Net with optimal thresholding for small blob detection in medical images. *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*. Canada. <https://doi.org/10.1109/coase.2019.8843234>
- Yen, J. C., Chang, F. J., & Chang, S. (1995). A new criterion for automatic multilevel thresholding. *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, 4(3), 370–378. <https://doi.org/10.1109/83.366472>
- Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., & Torr, P. H. S. (2015). Conditional Random Fields as Recurrent Neural Networks. *arXiv:1502.03240*. <https://doi.org/10.48550/arXiv.1502.03240>
- Zhu, Q., Dinh, T. H., Phung, M. D., & Ha, Q. P. (2021). Hierarchical convolutional neural network with feature preservation and autotuned thresholding for crack detection. *IEEE Access: Practical Innovations, Open Solutions*, 9, 60201–60214. <https://doi.org/10.1109/access.2021.3073921>