Binarization of degraded documents using a hybrid model between a classical method and a deep-learning approach

Milan Ćurković1 and Andrijana Ćurković2

1 University of Split, Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, Ruđera Boškovića 33, 21000 Split, Croatia

2 University of Split, Faculty of Science, Ruđera Boškovića 33, 21000 Split, Croatia

**Abstract.** Binarization of degraded documents is an important pre-processing step in character recognition that aims to separate the text from the background and irrelevant elements. Recently, deep learning methods have gained prominence over traditional approaches, although they often require extensive computational resources and training time. In this work, we propose a hybrid model that combines a classical binarization technique based on a parameterized 3D surface classification function with a deep learning approach. This integration reduces both the complexity of the model and the computation time compared to standard deep learning methods.

**Keywords:** Classic Method, Deep Learning Approach, Hybrid Model

1. Introduction

Classic degraded document binarization methods rely on basic parameters such as the global average and median, the local average and median, and the kernel size. In the end, they are part of a binarization decision function that integrates all parameters commonly accompanied by additional heuristic parameters. This approach mainly includes manually tunning heuristic parameters [1] or adaptively setting up some of the basic parameters [2–4] within the decision function defined based on experience. These problems have been largely overcome by modern CNN models [5, 6]. But like all optimization methods, they are not perfect and have certain shortcomings. They in general slower than classic methods, require labeled data for training, need diverse data for the training process for better generalization, and also have hyperparameters that need to be tuned (e.g. learning rate, layers). In this paper, we present a hybrid model between a classical method and a deep learning approach. All classical methods with their decision functions can be considered as a local optimum in minimizing the binarization process, where traditional measures such as FM (F-measure) or NRM (Negative Rate Metric) need to be optimized. The local minimum represents situations where one type of degraded document is well binarized while another is not.

1. Proposed method

The idea of the proposed approach is to use the CNN network model for building the approximation hypersurface of the arbitrary classical binarization model, including its decision function and heuristic parameters, to explore the space around the local minimum and make the necessary modifications to the approximation surface.

For the classical binarization model, we opted for a generic binarization method without manually tunable heuristic parameters [7]. Based on the calculation of the global text intensity $T\_{i}$, the global background intensity $B\_{i}$ and $m\_{tb}=max⁡(T\_{i}/255,1-B\_{i}/255)$, the method uses the logistic decision function

|  |  |
| --- | --- |
| $$fl\left(T\_{i}, B\_{i}\right)=m\_{tb}+\frac{1-m\_{tb}}{1+m\_{tb}^{2}\*e^{\left(m\_{tb}+\frac{{255}/{2}-T\_{i}}{m\_{tb}\*T\_{i}}\right)}}$$ | (1) |

in the final binarization decision function

|  |  |
| --- | --- |
| $$Bf\left(T\_{i}, B\_{i}\right)=\left\{\begin{matrix}1&I\left(x,y\right)<fl\*L\_{m} ∧fl<L\_{a} \\0&else\end{matrix}\right.$$ | (2) |

, where $L\_{m}$ represents the local median and $L\_{a}$ the local average values for pixel (x, y).

The following figure shows the function $fl$ as a depth field, where the abscissa axis is the background intensity $B\_{i}$, while the ordinate axis is text intensity $T\_{i}$.



**Fig. 1**. The logistic decision function as a depth field.

Since the method does not use heuristic parameters, the hypersurface we need to change in the CNN network model is the function $fl$. The idea of the full model is to approximate the $fl$ surface by a B-spline surface and use a neural network to adjust the B-spline surfaces to maximize FM and minimize NRM.

The proposed CNN model takes as input the original document patches (512×512 pixels, 1 channel), the current B-spline parameters (3 channels), and the target metrics FM and NRM (2 channels), and then predicts the adjustments of these parameters. The result is an updated B-spline surface that is used for thresholding. Behind the input data, the model uses precomputed median/average documents and ground truth binarization masks during the training process. The loss function is defined as follows:

|  |  |
| --- | --- |
| $$Loss =mse\left(FM, 1\right)+ mse\left(NRM, 0\right)$$ | (3) |

The actual binarization is still done using the traditional method, but with parameters that are learned by the neural network. The following figure shows the workflow of the proposed model.



**Fig. 2.** The proposed model workflow

The proposed hybrid model differs from standard neural networks in several key aspects. Unlike standard convolutional neural networks, which directly learn mappings from input images to binary outputs, the hybrid approach focuses on optimizing the parameters of a B-spline model embedded in a conventional binarization function. This integration combines the optimization of a neural network with a classical parametric model, whereas standard CNNs rely solely on learned features without explicit integration of such models. The hybrid framework can provide better interpretability as the B-spline parameters retain a clear, predefined role in the binarization process, as opposed to the less transparent decision making of conventional CNNs.

In terms of data efficiency, the hybrid model utilizes the prior knowledge encoded in the B-spline structure, which can reduce the amount of training data required compared to standard CNNs that need to completely relearn representations. Furthermore, the training goals differ significantly. The hybrid method uses loss functions derived from classical evaluation measures such as signal-to-noise ratio, F-measure or noise robustness to evaluate the quality of the binarized output. In contrast, standard CNNs often optimize pixel-wise losses such as binary cross entropy and focus on the direct correspondence between predicted and actual pixel values. These differences illustrate how the hybrid model combines classical image processing principles with modern learning-based techniques, balancing interpretability, data efficiency and the integration of domain knowledge with the purely data-driven end-to-end paradigm of standard neural networks.

The next figure shows the result of the proposed model after 1000, 1500 and 2000 epochs.

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |

**Fig. 3.** The result of surface modification. The first row shows the B-spline surface after 1000, 1500, and 2000 epochs; the second row shows its difference from the original B-spline surface.

1. Future work

Future directions include extending the hybrid approach to other classical methods with heuristic parameters and developing a supermodel that dynamically synthesizes multiple classical techniques. We will explore advanced parameterization strategies and validate the framework’s efficiency on constrained hardware.

Acknowledgments

This work was supported by the Croatian Science Foundation [HRZZ-IP-2022-10-8856]

References

1. Gatos B, Pratikakis I, Perantonis SJ (2006) Adaptive degraded document image binarization. Pattern Recognit 39:317–327

2. Singh TR, Roy S, Singh OI, Sinam T, Singh KM (2012) A New Local Adaptive Thresholding Technique in Binarization.

3. Vo GD, Park C (2018) Robust regression for image binarization under heavy noise and nonuniform background. Pattern Recognit 81:224–239

4. Chen X, Wang K, Wang Q (2012) An Adaptive Binarization Method for Camera Based Document Image. pp 677–684

5. 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

6. Pastor-Pellicer J, España-Boquera S, Zamora-Martínez F, Afzal MZ, Castro-Bleda MJ (2015) Insights on the Use of Convolutional Neural Networks for Document Image Binarization. pp 115–126

7. Ćurković M, Andrijana, Ćurković Binarization of degraded documents: A classic and generic method accompanied with a new semantic evaluation metric - MCHAM (May 20, 2025). Available at SSRN: https://ssrn.com/abstract=5262391. https://doi.org/10.2139/ssrn.5262391