Binarization of Degraded Documents Using a Hybrid Model Between a Classical Method and a Deep-Learning Approach

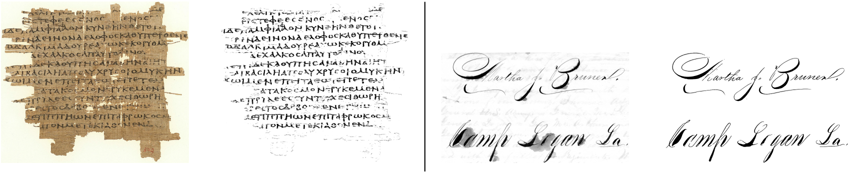
Milan Ćurković[1] Andrijana Ćurković[2]

[1]Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split, Croatia
[2]Faculty of Science, University of Split, Croatia

OES 2025

hrzz
Croatian Science Foundation

Degraded Documents: A Complex Problem



Classic Methods

- Global Thresholding (Otsu): Single threshold for entire image
- Local Thresholding (Sauvola): Pixel-wise adaptive thresholds
- Hybrid Thresholding (Our prior work [1]): Global thresholds and pixel-wise adaptive thresholds

Machine Learning Approaches

- Supervised: CNNs (U-Net) learn from labeled data
- Unsupervised: Clustering (K-means) for auto-segmentation
- GANs: Adversarial training for refined outputs

Binarization Approaches: Strengths & Limitations

Classic Methods

Advantages	Limitations
• Simple implementation • Fast execution (real-time)	• Manual parameter tuning required • Local optima convergence

Machine Learning Approaches

Advantages	Limitations
• Handles complex backgrounds • No manual parameter tuning • Adaptable to diverse input conditions	• Requires large labeled datasets • Overfitting risks • Limited generalization to unseen data • Computationally intensive - GPU needed

Key Trade-offs:

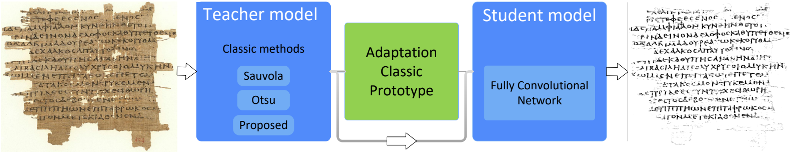
- Classic: Fast but inflexible
- ML: Robust but resource-intensive

Teacher-Student Framework

Core Idea: Combine classical methods' efficiency with ML's adaptability.

Teacher-Student Analogy:

- Teacher: Classical methods (Otsu, Sauvola, etc.) as prototype
  - Provides initial binarization rules
- Student: CNN model
  - Learns to enhance the teacher's output through deep optimization



Key Insight

Combine computational efficiency of classical methods with adaptive power of deep learning

Motivation

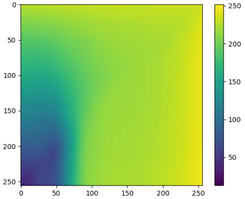
Overcome local optima in classical methods through parametric hypersurface optimization.

- Classical methods → local minima in optimization landscape
- CNN learns to approximate & adjust the binarization hypersurface

Classic binarization method [1]

Logistic decision function:

$$f(T_i, B_i) = m_{tb} + \frac{1 - m_{tb}}{1 + m_{tb}^2 * \exp\left(m_{tb} + \frac{255/2 - T_i}{m_{tb} * T_i}\right)} \quad (1)$$



Final binarization decision function:

$$\text{Binary}(x, y) = \begin{cases} 1 & \text{if } I(x, y) < f(T_i, B_i) \cdot L_m \wedge I(x, y) < L_a \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Text intensity T_i; Background intensity B_i; $m_{tb} = \max(T_i/255, 1 - B_i/255)$

B-spline Enhanced Hybrid Model

Hybrid Architecture:

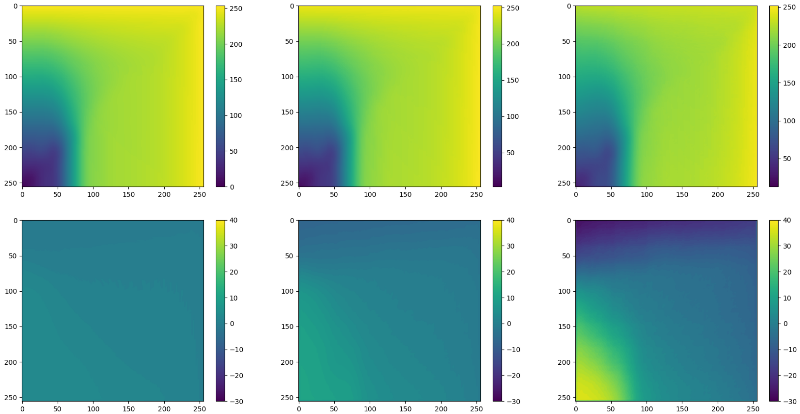
- Parametric Component: B-spline surface
- Neural Component: CNN adjusts B-spline parameters

Training Strategy:

- Per-pixel EMA (Exponential Moving Average): Selective grid updates
- Multi-term Loss: $\mathcal{L} = \alpha\mathcal{L}_{FM} + \beta\mathcal{L}_{NRM} + \gamma\mathcal{L}_{smooth}$
- Learnable temperature parameter

Binarization Formula:

$$\text{Output} = \sigma\left(\underbrace{\frac{(\text{Surface} \times \text{Median} - \text{Image})}{\text{Temperature}}}_{\text{B-spline term}} + \underbrace{\frac{(\text{Average} - \text{Image})}{\text{Temperature}}}_{\text{Classical term}}\right)$$

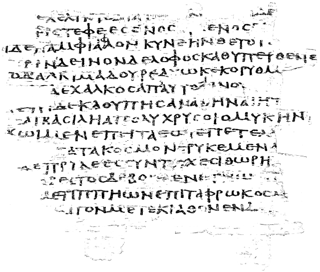


a) 500 epochs b) 1000 epochs c) 2000 epochs

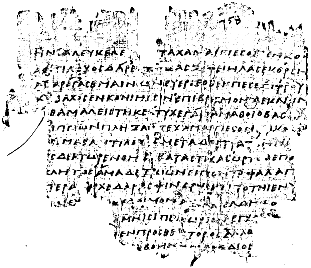
Table 1: Top row displays the B-spline surface updates after 500, 1000, and 2000 training epochs. Bottom row illustrates the differences between the original surface and each update at 500, 1000, and 2000 epochs.

- Hybrid: B-spline surface after 2000 epochs

Dataset	FM		NRM	
	Original	Hybrid	Original	Hybrid
DIBC02009	0.853428	0.857464	0.098911	0.098286
DIBC02011	0.852309	0.852111	0.105821	0.104770
DIBC02013	0.854460	0.855951	0.102862	0.100566
DIBC02016	0.886066	0.887100	0.080498	0.080044
DIBC02018	0.715489	0.714246	0.152436	0.153628
DIBC02019	0.727425	0.734166	0.113878	0.113998



Original



Hybrid

Enhanced Hybrid Architectures - Future Work

Two U-Net Models Incorporating the Proposed Classical Method [1]

- **Standard U-Net Model** – Utilizes a conventional U-Net architecture to directly predict the binarized output from the input features, including the initial binarization.
- **Hybrid U-Net Model** – Integrates the classical binarization formula to compute an initial binarization, which is then provided as an additional input channel to the U-Net.

	Standard U-Net model	**Hybrid U-Net model**
Input	4 channels: Image, Median, Average, B-spline	5 channels: Adds classical binarization output
Architecture	Standard U-Net	Classical binarization $\rightarrow$ U-Net $\rightarrow$ Residual blend
Output	Direct prediction	$\alpha \times$ Refined $+ (1 - \alpha) \times$ Classical
Focus	Pure deep learning-based segmentation	Incorporates classical methods for robustness
Adaptability	Limited to learned features	Leverages prior domain knowledge

Conclusion

Key Contributions

- Novel hybrid binarization framework
- Teacher-student knowledge transfer
- B-spline surface optimization
- Multi-term loss function

Practical Impact

- **Cultural heritage**: Digitization of ancient manuscripts
- **Legal**: Archival document processing
- **Education**: Historical text preservation

Thank You!

Questions?
milan.curkovic@fesb.hr

Croatian Science Foundation

References

[1] Milan Ćurković and 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*. DOI: 10.2139/ssrn.5262391. URL: https://www.ssrn.com/abstract=5262391.