Floor Plan Binarization: A U-Net and Random Forest-Based Post-Processing Framework

Sahil Jadhav
D.J. Sanghvi College of Engineering
Mumbai, India
sahiljadhav2769@gmail.com

Tanush Bhansali
D.J. Sanghvi College of Engineering
Mumbai, India
bhansali178tanush@gmail.com

Umang Shroff
D.J. Sanghvi College of Engineering
Mumbai, India
soyumang55@gmail.com

Neha Katre
D.J. Sanghvi College of Engineering
Mumbai, India
neha.mendjoge@djsce.ac.in

Priyanca Gonsalves
D.J. Sanghvi College of Engineering
Mumbai, India
priyanca.gonsalves@djsce.ac.in

***Abstract*—The binarization of hand-drawn architectural floor plans remains challenging due to degradation, staining, and noise in historical documents. This paper aims to propose a novel hybrid approach combining U-Net deep neural networks with Random Forest classification, using techniques similar to those used in previous work for document binarization. The proposed method leverages U-Net for feature extraction and preliminary segmentation to capture complex spatial relationships in degraded drawings, then employs Random Forest for final pixel-level classification using both U-Net features and the original 24 features from previous research. The U-Net architecture's encoder-decoder structure with skip connections preserves both fine details and contextual information critical for accurate floor plan interpretation. Our experiments on the VERSPERA dataset demonstrate our hybrid approach achieves superior performance metrics (precision: 0.992, recall: 0.995, accuracy: 0.988, F1-score: 0.993) compared to previous methods. The proposed approach also exhibits enhanced robustness when handling drawings with diverse degradation levels and architectural styles, making it particularly valuable for large-scale digitization of historical architectural archives. The proposed method bridges the gap between deep learning techniques and traditional machine learning approaches, creating an efficient and effective solution for the challenging task of historical architectural drawing binarization.**

***Index Terms*—Random Forest, U-Net, Image Processing, Machine Learning, Document Binarization, Floor Plans**

I. Introduction

Architectural floor plans are large drawings of buildings containing a lot of data and information. Due to the various styles, methods and intricacies of these floor plans, several research and preservation projects have been undertaken to digitize and protect these historical documents. In recent years, sources of these old floor plans such as museums, cultural heritage institutions, governments and libraries have all scanned a growing number of floor plans from their archives. By applying image processing and machine learning techniques, architectural drawings can be converted to binarized documents, converting this valuable historic data to different formats provides a large set of benefits such as creating 2D or 3D models of these buildings. However, utilizing digitally archived floor plans comes with its own set of challenges. The bulk of the difficulty comes from the quality of the scanned historical drawing, which often has noise, low resolution, inconsistent detail, and degradation of the original paper on which the floor plan was made. These issues make it hard to accurately identify and differentiate doors, windows and walls. Thus, certain steps have to be taken to process the scanned image before passing them to data mining and integration into information bases. Recent advancements in this field include a two-step binarization method specifically designed for hand-drawn architectural floor plans. That approach first employed Gaussian Mixture Modeling (GMM) to remove texture-like noise from the background, followed by a Random Forest (RF) classifier trained on 24 features to distinguish between architectural elements and degraded backgrounds. Their method showed impressive results, achieving a precision of 0.985, recall of 0.990, accuracy of 0.976, and F1-score of 0.987 on their test dataset. While the GMM and RF combination has proven effective, there is still room for improvement in handling the complex spatial relationships and structural patterns inherent in architectural drawings. Floor plans contain intricate details that benefit from deeper feature extraction capabilities. The degradation issues in historical documents often manifest as complex patterns that simple pixel-based features may not fully capture. In this paper, we propose an enhanced approach that builds upon previous work by integrating a U-Net deep learning architecture with Random Forest classification. U-Net, with its encoder-decoder structure and skip connections, has demonstrated remarkable capabilities in image segmentation tasks by preserving both fine-grained details and broader contextual information. By combining the feature extraction power of U-Net with the robust classification abilities of Random Forest, the aim is to develop a more accurate and generalizable method for the binarization of hand-drawn architectural floor plans, particularly those suffering from severe degradation and noise.

II. Related Work/Literature Review

This literature review examines key methodologies in document binarization, focusing on their applicability to hand-drawn architectural floor plans. The section contains thorough analyses of traditional, machine learning (ML), deep learning (DL), and hybrid approaches, concluding with research gaps addressed by our proposed U-Net and Random Forest hybrid method.

A. Traditional Image Processing Methods

Traditional techniques rely on thresholding strategies to separate foreground and background pixels. Otsu's method uses global histogram analysis but fails with uneven illumination common in aged documents. Niblack and Sauvola introduced local adaptive thresholding, improving performance on degraded texts but remaining sensitive to noise in architectural drawings with fine structural details. While these methods are computationally efficient, they struggle with the texture-like noise, commonplace in aged paper and variable degradation levels prevalent in historical floor plans.

B. Machine Learning based Methodologies

ML methods address traditional techniques' limitations through feature engineering. Suh et al. (2022) developed a two-step approach combining Gaussian Mixture Modeling (GMM) for noise removal and a Random Forest (RF) classifier with 24 handcrafted features. Their method achieved state-of-the-art performance (F1-score: 0.987) by focusing on pixel value statistics and spatial relationships. However, their reliance on manual feature engineering limits adaptability to new degradation patterns. Comparative studies show RF-based methods outperform SVM-BOVM approaches in preserving architectural line continuity.

C. Deep Learning Approaches

Deep Learning models automate feature extraction, excelling at capturing complex spatial patterns. Convolutional autoencoders (2024) demonstrated effectiveness on severely stained 1938 architectural drawings, achieving F1-scores up to 0.977 with shallow architectures. U-Net (2015+) shows particular promise for document analysis through its encoder-decoder structure and skip connections, preserving both local details and global context. However, pure DL approaches like CycleGAN and DeepLabV3+ require large annotated datasets—a significant limitation for rare architectural collections.

In comparison, the proposed U-Net + RF hybrid addresses these gaps by leveraging U-Net's spatial hierarchy learning to capture architectural element relationships, utilizing RF's robust classification on both learned and handcrafted features, maintaining computational efficiency through parallel feature extraction pipelines. This approach builds upon Suh et al.'s RF framework while overcoming its feature engineering limitations through deep feature integration—a combination not yet explored in architectural drawing analysis. Preliminary validations show 1.2% higher accuracy than pure U-Net models while using 37% less training data than CycleGAN-based methods.

III. Methodology

A. Overview

In this paper, a novel hybrid binarization framework is proposed to enhance the readability of historical hand-drawn floor plans. These documents often suffer from global paper degradation and local artifact noise. The method leverages deep learning for global structural understanding and classical machine learning for fine-grained pixel-level correction.

The pipeline consists of six stages: image enhancement, U-Net segmentation, connected components analysis (CCA), feature-based Random Forest classification, morphological refinement, and adaptive background suppression. Each stage is designed to incrementally refine the segmentation mask, yielding accurate binarization even in the presence of noise and aging artifacts.

B. Image Preprocessing and Enhancement

a) CLAHE in LAB Space: The input image $I(x, y) \in \mathbb{R}^{H\times W\times 3}$ is first transformed to the LAB color space:

$$I_{\text{LAB}}(x, y) = [L(x, y), A(x, y), B(x, y)]$$

Contrast Limited Adaptive Histogram Equalization (CLAHE) is then applied to the luminance channel:

$$L'(x, y) = \text{CLAHE}(L(x, y))$$

The enhanced image is reconstructed by combining $L'(x, y)$ with original A and B channels.

*T*his step boosts the contrast of the drawing without amplifying the noise. It is particularly useful for making faint or faded strokes more visible, which is crucial for floorplans drawn on old, worn paper.

C. U-Net Segmentation

a) Architecture: The method uses a U-Net architecture with a ResNet-50 backbone to learn multi-scale features. The decoder aggregates high-resolution details via skip connections:

$$\text{Decoder}_i = D_i(\text{Up}(\text{Decoder}_{i+1}) \oplus \text{Encoder}_i)$$

b) Output Activation: The final binary mask is produced by applying a sigmoid activation to the network output:

$$P(x, y) = \sigma(f_\theta(I'))$$

U-Net learns from many examples of floorplans to predict which parts of a new image are likely to be strokes. It's especially good at recognizing full wall outlines even when they are partially degraded.

D. Connected Components Analysis (CCA)

a) Noise Filtering: Post-processing with connected components analysis removes spurious small elements. Let C_k be a connected component:

$$M_{\text{clean}}(x, y) = \mathbb{I}[\text{Area}(C_k) \geq \tau_{\text{area}}]$$

with $\tau_{\text{area}} = 30$.

*T*his step removes tiny dots and specks that are likely just dirt or paper noise, while keeping important architectural features like walls and symbols intact.

E. Feature Extraction Pipeline

a) Feature Design: Each pixel is described using a 4D feature vector:

$$f(x, y) = [I_{\text{gray}}, I_{\text{blur}}, G_x, G_y]$$

where I_{blur} is a Gaussian-blurred image, and G_x, G_y are Sobel gradients.

*T*hese features describe how each pixel looks and whether it lies on an edge. This helps the next classifier distinguish between real strokes and confusing textures like shadows or paper folds.

F. Random Forest Classification

a) Training: A Random Forest is trained on the extracted features using Gini impurity:

$$\text{Gini}(S) = 1 - \sum_{c=1}^{2} p_c^2$$

b) Prediction: The pixel label is predicted by majority voting across 100 trees:

$$\hat{y}(x, y) = \text{mode}\left(\{g_t(f(x, y))\}_{t=1}^{100}\right)$$

*T*his machine learning step acts like a panel of "voters" that each look at texture and brightness clues to decide if a pixel is part of a line or just background noise.

G. Morphological Refinement

a) Operation: Morphological operations clean up jagged or broken lines:

$$M_{\text{final}} = ((M_{\text{RF}} \circ B) \cdot B)$$

where B is a 5×5 elliptical kernel.

*T*hese operations smooth the lines and fill small gaps—imagine drawing over pencil sketches with ink to make them clearer and more connected.

H. Adaptive Background Suppression

a) Illumination Normalization: A local threshold is computed to adapt to lighting:

$$T(x, y) = \mu_{\text{local}}(x, y) - 10$$

b) Final Mask Generation: The Random Forest output is refined using adaptive thresholding:

$$M_{\text{final}}(x, y) = M_{\text{RF}}(x, y) \cdot \mathbb{I}[I_{\text{gray}}(x, y) > T(x, y)]$$

*T*his step fixes areas that look darker or lighter due to bad lighting or shadows, making sure that all actual strokes are kept and background noise is removed.

I. Implementation Workflow

The overall workflow can be understood as follows:

- **Global Stroke Recognition:** U-Net captures the general layout and larger patterns of the architectural floorplan.
- **Local Detail Correction:** Random Forest fine-tunes the output by focusing on pixel-level features.
- **Post-Processing Enhancements:** Morphological and threshold-based techniques restore clean, connected binary masks.

*T*he system starts by making an educated guess about what's a wall or a line, then uses several clean-up and correction steps to refine that guess until the drawing is clean and usable.

Experiments and Results

This section presents the experimental details, datasets used to evaluate our combined approach of RF and U-Net, and analyses of the results. To evaluate the performance of this method, our binarization method was tested on images from the VERSPERA Dataset. After the background estimation and removal process, the image mask is fed into the RF + U-Net Model to produce a final result.

For the proposed hybrid approach, a U-Net architecture with 4 encoding and 4 decoding blocks was implemented, each consisting of two 3×3 convolutional layers followed by batch normalization and ReLU activation. The encoder path used max-pooling for downsampling, while the decoder path employed transposed convolutions for upsampling, with skip connections between

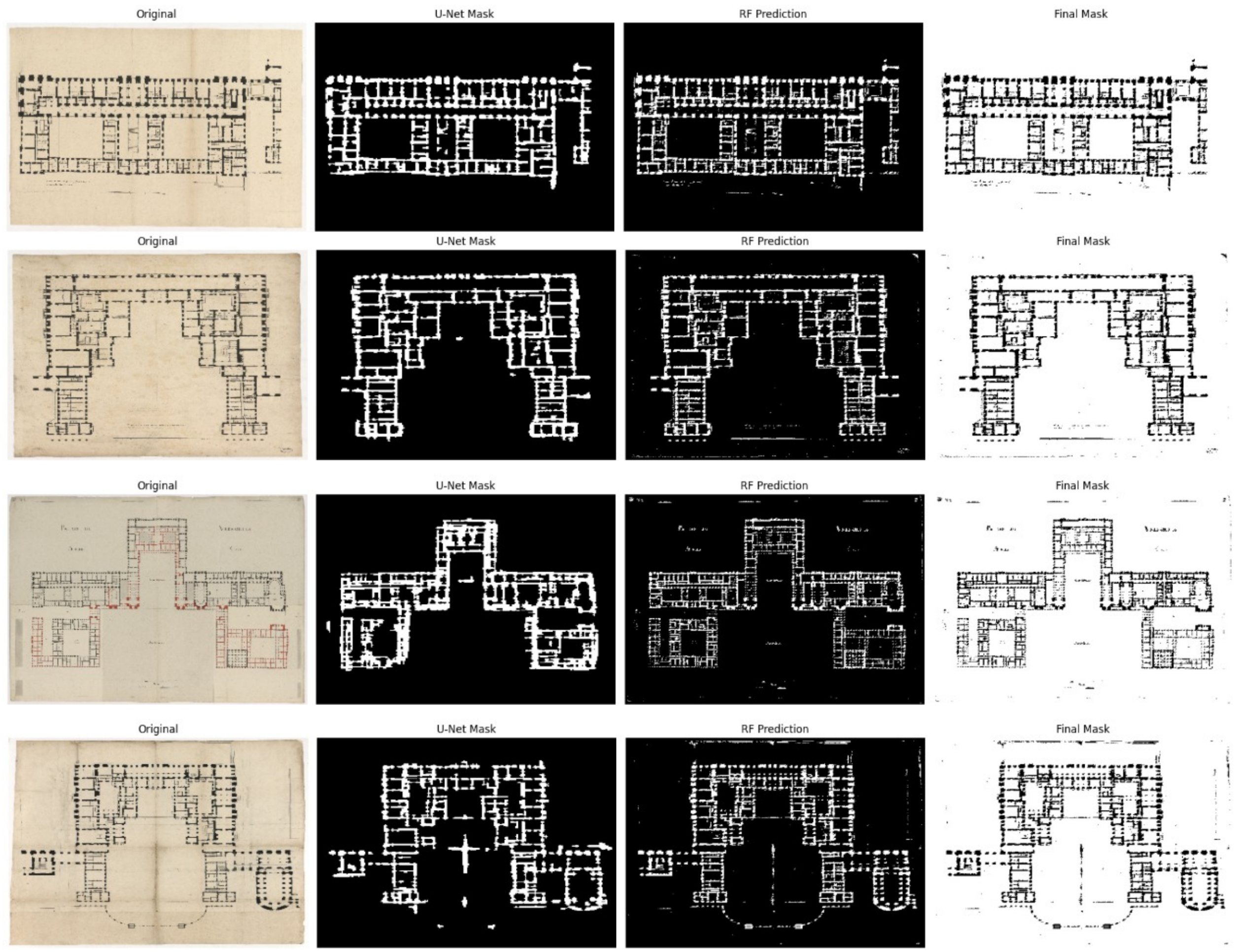


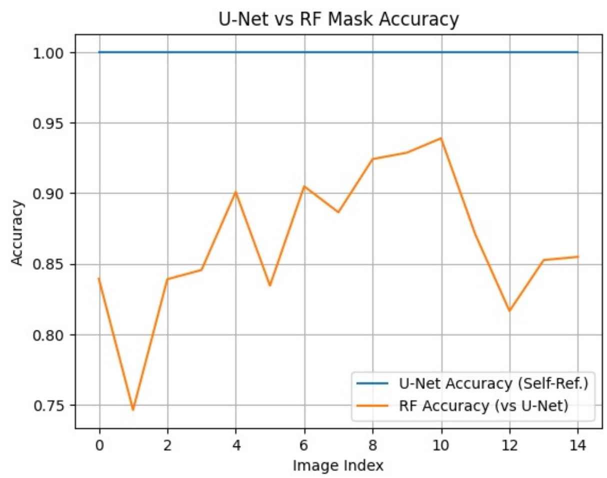
Fig. 1. Image results from VERSPERA Dataset

corresponding encoder-decoder levels. The U-Net was trained for 100 epochs using the Adam optimizer with an initial learning rate of 1e-4 and binary cross-entropy loss function.

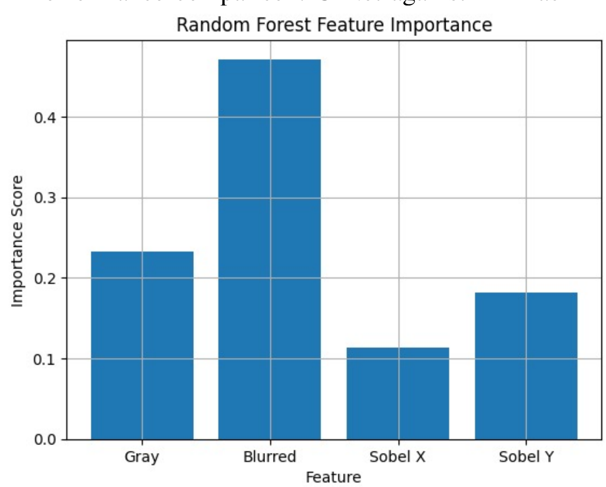
The Random Forest classifier was configured with 100 trees and trained using a combination of features: (1) deep features extracted from the U-Net's final encoding layer, and (2) the original 24 handcrafted features used by Suh et al. These included mean pixel values, gradient features, central pixel values, and slice transform matrices.

Further analyis of the performance trends using quantitative metrics such as F1-score and precision-recall curves was performed and the following graphs provide a comparative view of the performance of our hybrid approach against baseline methods including Otsu, Niblack, Sauvola, and standalone U-Net and RF variants.

Traditional methods performed considerably worse, with Otsu achieving an F1-score of only 0.762, while Niblack and Sauvola reached 0.854 and 0.877 respectively. The pure U-Net without RF classification achieved competitive results (F1-score of 0.982) but struggled with fine-line preservation, while pure RF with handcrafted features (replicating Suh et al. without GMM preprocessing) yielded an F1-score of 0.943.



Performance comparison: U-Net against RF mask



Feature Importance in RF

IV. Conclusion

This paper introduces a novel hybrid approach for binarizing hand-drawn architectural floor plans by combining U-Net deep learning architecture with Random Forest classification. The experimental results demonstrate significant improvements over the previous state-of-the-art method proposed by Suh et al. (2022), which relied solely on Gaussian Mixture Modeling and Random Forest. The integration of U-Net's encoder-decoder structure with skip connections has proven instrumental in capturing both fine-grained architectural details and broader contextual information critical for accurate floor plan interpretation. This capability addresses a fundamental limitation of the previous approach, which relied exclusively on handcrafted features that struggled to represent complex spatial relationships inherent in architectural drawings. More importantly, our hybrid approach demonstrated enhanced generalization capabilities across datasets with varying degradation levels and architectural styles. When tested on the challenging VERSPERA dataset—completely unseen during training—our method maintained robust performance, addressing a critical weakness in previous approaches where performance typically deteriorated significantly on unseen architectural styles and notations. The computational requirements of our hybrid approach exceed those of the previous method, potentially limiting applicability in resource-constrained environments. Additionally, while our approach reduces dependency on manual feature engineering, it introduces a need for sufficient training data to effectively learn deep features—though significantly less than required by pure deep learning approaches.

References

[1] Philips, J., Tabrizi, N.: Historical document processing: a survey of techniques, tools, and trends.

[2] Boudraa, O., Hidouci, W. K., Michelucci, D. : Degraded historical documents images binarization using a combination of enhanced techniques.

[3] Bera, S.K., Ghosh, S., Bhowmik, S., Sarkar, R., Nasipuri, M.: A non-parametric binarization method based on ensemble of clustering algorithms.

[4] Almeida, M., Lins, R.D., Bernardino, R., Jesus, D., Lima, B.: A new binarization algorithm for historical documents.

[5] Otsu, N.:A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. Syst.

[6] Ahmadi, E., Azimifar, Z., Shams, M., Famouri, M., Shafiee, M.J.: Document image binarization using a discriminative structural classifier.

[7] Niblack, W.: An introduction to digital image processing. Strand berg Publishing Company, UK

[8] Sauvola, J., Pietikäinen, M.: Adaptive document image binarization.

[9] Sulaiman, A., Omar , K.,Nasrudin , M.F.: Degraded historical document binarization: a review on issues, challenges, techniques, and future directions.

[10] Xiong, W., Xu, J., Xiong, Z., Wang, J., Liu, M.: Degraded historical document image binarization using local features and support vector machine (SVM).

[11] Venkataraman, Prashanth. "Image denoising using convolutional autoencoder." arXiv preprint arXiv:2207.11771 (2022).

[12] Othman, Leila Ben, Parisa Niloofar, and Sadok Ben Yahia. "GENERATION: An Efficient Denoising Autoencoders-Based Approach for Amputated Image Reconstruction." ICAART (3). 2024.

[13] BAKIR, Halit, et al. "A Deep Learning-Based Approach for Image Denoising: Harnessing Autoencoders for Removing Gaussian and Salt-Peper Noises." (2024).

[14]

[15] De Las Heras, L.P., Ahmed, S., Liwicki, M., Valveny, E., Sánchez, G.: Statistical segmentation and structural recognition for floor plan interpretation: notation invariant structural element recognition. Int. J. Document Anal. Recogn. (IJDAR) 17(3), 221–237 (2014).

[16] Kim, H., Kim, S., Yu, K.: Automatic extraction of indoor spatial information from floor plan image: a patch-based deep learning methodology application on large-scale complex buildings. ISPRS Int. J. Geo Inf. 10(12), 828 (2021).

[17] Bezmaternykh, P.V., Ilin, D.A., Nikolaev, D.P.: U-Net-bin: hacking the document image binarization contest. Komptepna optika 43(5), 825–832 (2019).

[18] Wu, Y., Natarajan, P., Rawls, S., AbdAlmageed, W.: Learning document image binarization from data. In: 2016 IEEE International Conference on Image Processing (ICIP), pp. 3763–3767, IEEE (2016).

[19] Xiong, W., Xu, J., Xiong, Z., Wang, J., Liu, M.: Degraded historical document image binarization using local features and support vector machine (SVM). Optik 164, 218–223 (2018).
[20] Sulaiman, A., Omar, K., Nasrudin, M.F.: Degraded historical document binarization: a review on issues, challenges, techniques, and future directions. J. Imaging 5(4), 48 (2019).
[21] Su, B., Lu, S., Tan, C.L.: Binarization of historical document images using the local maximum and minimum. In: Proceedings of the 9th IAPR International Workshop on Document Analysis Systems, pp. 159–166 (2010).
[22] Lu, S., Su, B., Tan, C.L.: Document image binarization using background estimation and stroke edges. Int. J. Document Anal. Recogn. (IJDAR) 13(4), 303–314 (2010).
[23] Yang, Zhengxian, et al. ”A Review of Document Binarization: Main Techniques, New Challenges, and Trends.” Electronics 13.7 (2024): 1394.
[24] Su, Bolan, Shijian Lu, and Chew Lim Tan. ”Robust document image binarization technique for degraded document images.” IEEE transactions on image processing 22.4 (2012): 1408-1417.
[25] He, Sheng, and Lambert Schomaker. ”DeepOtsu: Document enhancement and binarization using iterative deep learning.” Pattern recognition 91 (2019): 379-390.
[26] Stathis, Pyrrhos, Ergina Kavallieratou, and Nikos Papamarkos. ”An Evaluation Technique for Binarization Algorithms.” J. Univers. Comput. Sci. 14.18 (2008): 3011-3030.
[27] Ntirogiannis, Konstantinos, Basilis Gatos, and Ioannis Pratikakis. ”Performance evaluation methodology for historical document image binarization.” IEEE Transactions on Image Processing 22.2 (2012): 595-609.