Scientific Journal of Informatics
Vol. 13, No. 1, February 2026
p-ISSN 2407-7658 https://journal.unnes.ac.id/journals/sji/index e-ISSN 2460-0040

Performance Evaluation of Otsu and Sauvola Thresholding for Structured Document Binarization

Muhammad Noko Darpito[1]*, Kartika Firdausy[2], Abdul Fadlil[3]
[1]Master Program of Informatics, Faculty of Industrial Technology, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[2,3]Department of Electrical Engineering, Faculty of Industrial Technology, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

Abstract.
Purpose: Digitizing public administration records, particularly structured forms such as the Transport of Plants and Wildlife Abroad (Surat Angkut Tumbuhan dan Satwa Liar Luar Negeri / SATS-LN), necessitates meticulous preparation for precise subsequent analysis. Most of the photos in the SATS-LN archives are scanned, and they have inconsistent lighting, varying resolution, and background noise, which makes it difficult to separate the text from the backdrop and read it clearly. This work identifies the optimal SATS-LN binarization approach for preserving textual structure and suppressing background artifacts.
Methods: A four-stage pipeline is used. First, Detectron2 localizes seven important SATS-LN fields. Second, binarization is investigated with global Otsu and adaptive Sauvola thresholding under three parameter configurations. Third, following binarization, Contrast-Limited Adaptive Histogram Equalization (CLAHE) boosts local contrast. Finally, Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), Difference from Reference for Distortion (DRD), Precision, Recall, F1-score, and Foreground Ratio are assessed on 200 annotated SATS-LN documents (150 scanner-based/DOC and 50 camera captured/CAM).
Result: The acquisition domain and assessment model affect binarization performance on 200 SATS-LN documents (150 DOC scans and 50 CAM images). Global Otsu_T10 has the highest median PSNR (21.19 dB) and the lowest median MSE (494.69), indicating a visually cleaner background. However, segmentation-based metrics show better stroke preservation with Sauvola, as Sauvola_k05 has the strongest DOC text–background separation (F1 = 0.938). In the CAM domain, where illumination variability dominates, Sauvola performs better across structural and segmentation indicators, with Sauvola_k04 performing best overall (F1 = 0.980) and mitigating the over-segmentation tendency of strict global thresholds. The Sauvola window (25x25) and CLAHE clip limit (1.0) results suggest using Sauvola_k05 for DOC and Sauvola_k04 for CAM to preserve text integrity and reduce background artifacts.
Novelty: This study presents a novel field-level binarization assessment that combines automated cropping and ground-truth evaluation, providing practical guidance for robust preprocessing that supports scalable, reliable, and cross-device public document digitization.

Keywords: Image thresholding, Otsu, Sauvola, Clahe, Document scanned
Received January 2026 / **Revised** February 2026 / **Accepted** February 2026

INTRODUCTION

Government agencies increasingly transition from conventional archival practices to digital document management to improve accessibility, transparency, and processing efficiency [1]. The Transport of Plants and Wildlife Abroad (SATS-LN) documents are official, structured forms used for licensing the export of natural plants and wildlife. They contain multiple fixed-position elements (e.g., document number, exporter and destination addresses, validity period, origin/destination ports, and a table of listed items), making them suitable for layout-based information extraction [2].

In practice, many SATS-LN archives are stored as scanned images or photographs. Visual quality differs substantially due to acquisition resolution, illumination, shadows, paper stains, and background non-uniformity. These issues reduce text readability and complicate subsequent analysis stages, especially when text strokes are thin or partially degraded [3], [4]. Therefore, robust preprocessing is required before downstream extraction.

[1]*Corresponding author.
Email addresses: 2407048001@webmail.uad.ac.id (Darpito)
DOI: 10.15294/sji.v13i1.40245

Recent research has highlighted the importance of adaptive and learning-based binarization techniques in addressing illumination and degradation issues across various document sources. Dey and Jawanpuria (2022) propose a confidence-based enhancement of Sauvola's technique to improve the accuracy of unsupervised foreground-background separation on diverse document datasets [5]. Asatryan et al. (2023) conducted a comparison of Otsu and Sauvola thresholding for historical handwritten texts, indicating that local adaptive approaches enhanced readability under fluctuating backdrop conditions [6]. Wibawa and Anggraeni (2023) discovered that the Otsu and Sauvola binarization algorithms yielded similar Optical Character Recognition (OCR) accuracy across diverse document noise levels, with Sauvola exhibiting marginally superior performance under non-uniform lighting [7].

Naik (2024) also demonstrated that combining CLAHE and Sauvola binarization significantly improved the readability of degraded documents and yielded higher quantitative scores, such asSNR and DRD [3]. A thorough analysis by Bataineh and Tounsi (2025) reaffirmed the enduring significance of threshold-based and hybrid enhancement techniques, highlighting their versatility in both scanner and camera-based document contexts [8]. These studies collectively demonstrate that conventional thresholding techniques, when combined with localized contrast normalization methods such as CLAHE, remain competitive with contemporary deep learning algorithms for document binarization tasks.

However, most of the studies that have been done so far only look at how well binarization works on the whole document or focus on OCR accuracy as the main result. They don't look at how binarization quality changes from one field to another in structured documents. Also, there aren't many systematic comparisons between documents scanned and those taken with a camera that use the same cropped areas and the same binary ground truth. As a result, there is no practical, evidence-based advice on how to choose the best binarization strategies for different types of acquisition domains and structured document layouts.

Binarization is an important step in preprocessing. It changes grayscale areas of documents into binary foreground and background representations. Global approaches, such as Otsu, work well on scans with even lighting, while adaptive local methods, like Sauvola, are designed to work with uneven lighting by using local statistics to set thresholds [9], [10]. CLAHE can also adjust for changes in lighting and improve the visual uniformity across regions. This is especially important for papers taken with a camera [11], [12].

This study compares Otsu and Sauvola thresholding (each with three parameter configurations) in conjunction with CLAHE, assessing their performance across two acquisition domains (DOC vs. CAM). The contribution is an unbiased, field-level assessment of visual-similarity metrics (PSNR, SSIM) and segmentation-accuracy metrics (Precision, Recall, F1, DRD, FG Ratio) in comparison to binary ground truth.

METHODS

Before quantitative assessment, each SATS-LN document undergoes preprocessing. First, Detectron2 detects and crops relevant text portions in the image to focus on specific areas. Each cropped region is binarized using global Otsu and adaptive local Sauvola thresholding methods to create several binary representations of the same content. After binarizing the output, CLAHE enhances local contrast and distinguishes foreground text strokes from the background, resulting in a normalized image suitable for objective image-quality evaluation. The overall workflow is presented in Figure 1.

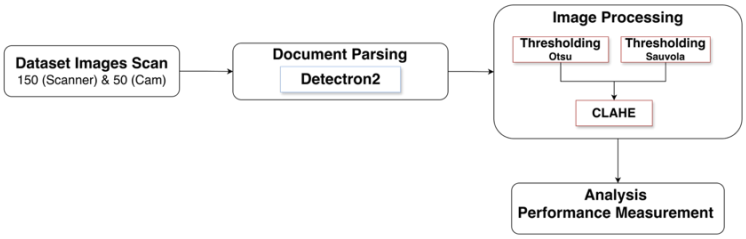


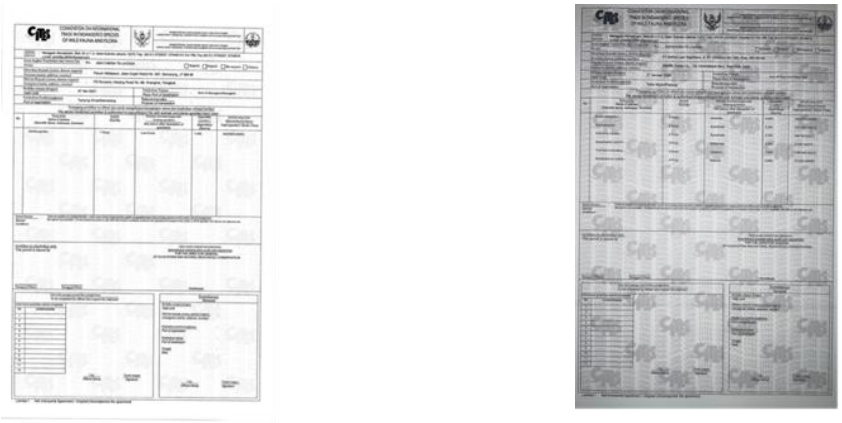
Figure 1. Overview of the experimental methodology

Dataset and Document Domain

The dataset used in this study consists of documents related to the Transport of Plants and Wildlife Abroad (SATS-LN), which have a fixed structure with data elements such as document number, name and address

of exporter, name and address of destination, validity period, port of origin, port of destination, and a table listing the types of goods.

The dataset contains two acquisition domains: DOC and CAM. The DOC domain has 150 scanned SATS-LN photos in PNG format, each with a resolution of 2550 × 3893 pixels. This ensures a stable visual quality, with minimal geometric distortion and a uniform backdrop appearance. There are 50 camera-captured photos in the CAM domain, each with a resolution of 2266 × 3553 pixels. These images show changes in lighting. Figures 2 (a) DOC (scanner-based image) and (b) CAM (camera-captured image) display SATS-LN document samples from two acquisition zones.



(a) DOC (b) CAM

Figure 2. Representative SATS-LN document samples from two acquisition domains

Using bounding boxes, all elements are labeled in COCO JSON format, which ensures that the spatial labeling is consistent across all texts. Detectron2 uses these annotations to learn and make predictions, which result in cropped pictures for each field that are used in the next steps of binarization and evaluation.

Document Element Detection Using Detectron

To isolate relevant regions and reduce background variability, object detection is applied to each SATS-LN document using Detectron2 with a Faster Region-based Convolutional Neural Network (R-CNN) architecture and a Residual Network with 50 layers (ResNet-50) Feature Pyramid Network (FPN) backbone [13], [14], [15], [16]. Faster R-CNN combines a Region Proposal Network (RPN) to generate candidate regions, followed by equivalent classification and bounding-box regression to refine labels and localization [17], [18], [19] ResNet-50 provides stable feature extraction via residual connections, and FPN enhances multi-scale detection by fusing high-level semantic and low-level spatial features [20], [21]. The trained model outputs bounding boxes for seven classes, and each box is cropped and saved as a per-element image; using identical crops across all binarization scenarios ensures that metric differences reflect thresholding performance rather than region selection.

Image Binarization (Otsu and Sauvola Thresholding)

The goal of the binarization stage is to convert the cropped image into a two-level grayscale image (black and white), allowing the text area to stand out clearly against the backdrop. This study employs two primary methodologies, specifically Otsu and Sauvola, which exemplify global and local adaptive thresholding techniques.

Otsu method

Otsu selects a global threshold that maximizes between-class variance between background and foreground intensities [22], [23], [24]. In theory, this approach is the same as maximizing the variance between classes, which is defined in Equation 1.

$$\sigma_B^{2(T)} = \omega_0\omega_1(\mu_0 - \mu_1)^2 \quad (1)$$

Here, ω_0 and ω_1 are the chances of the background and foreground classes and μ_0 and μ_1 is the average intensity of each class [25]. In this study, the Otsu threshold is adjusted to assess sensitivity, as shown in Equation 2.

$$T = \alpha \times T_{\text{Otsu}},\ \alpha \in \{0.9,\ 1.0,\ 1.1\} \tag{2}$$

As shown in Equation 3, clipping ensures that the scaled threshold remains within the valid 8-bit range.

$$T_{adj} = \text{clip}\ (\alpha \cdot T_{\text{Otsu}},\ 1,\ 254) \tag{3}$$

Clip restricts the threshold value to avoid degenerate binarization when $T$is too close to 0 or 255. Finally, the binarization rule is defined as: if $I(x,y) > T_{adj}$, the pixel is assigned 255 (background), otherwise it is assigned 0 (foreground).

Sauvola method

This method improves upon Niblack's approach by incorporating both local contrast and standard deviation normalization, which allows it to perform better in documents with uneven illumination and noise. Sauvola's adaptive strategy dynamically adjusts thresholds for each region, leading to more robust text-background separation, especially in degraded or historical document images [26], [3].

The Sauvola method builds on the Niblack method by utilizing the intensity statistics of pixels surrounding the observation area to determine the threshold value locally. Equation 4 shows the Sauvola threshold formula.

$$T(x,y) = m(x,y)\left[1 + k\left(\frac{s(x,y)}{R} - 1\right)\right] \tag{4}$$

where $m(x,y)$ is the local mean, $s(x,y)$ is the local standard deviation, R is the maximum standard deviation (typically R = 128 for 8-bit images), and k is the sensitivity control parameter.

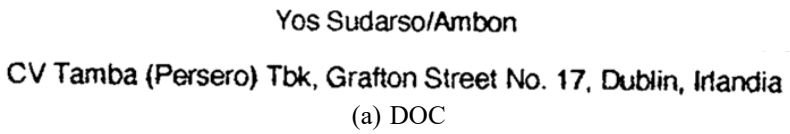
In this study, k is changed from 0.3 to 0.4 to 0.5 to see how adaptivity affects the results of binarization, the clarity of the text, and the background noise. If you use smaller k values, the background will be cleaner, but the text will be thinner. If you use larger k values, the text will be bolder, but the background noise will be louder.

Contrast Enhancement Using CLAHE

CLAHE is used after thresholding to fix problems with local contrast and uneven lighting. CLAHE performs adaptive histogram equalization on small tiles (8×8) with a clip limit of 1.0 to prevent noise from becoming too loud. Using CLAHE consistently across different situations makes it easier to compare fairly in the same contrast-normalized condition, especially for CAM photos with shadows and uneven lighting [27], [28] [29].

Ground Truth Construction

For objective segmentation evaluation, each cropped element is paired with a manually created binary reference image. Ground truth is generated using the PhotoScape X application by adjusting brightness and contrast and manually cleaning background artifacts while preserving stroke boundaries, then saved as a PNG file with the same resolution as the crop to provide an ideal text–background separation for metric computation [30], [31]. Figure 3 (a) shows a binary ground truth example from the DOC domain (scanner-based crop). The lighting is mostly even, and manual refinement mostly eliminates leftover background shading while preserving character strokes. Figure 3 (b) shows the same ground truth from the CAM domain (camera-captured crop). This one usually requires more careful cleaning because the lighting and contrast aren't always consistent, so strokes remain intact, and background artifacts are removed.



(a) DOC

(b) CAM

Figure 3. Binary ground-truth image after manual refinement using PhotoScape X

Evaluation Metrics

This study employs two evaluation models: visual similarity between the original crop and the binarized image, and segmentation accuracy between the binarized image and the binary ground truth.

Visual quality evaluation

Visual quality analysis, typically evaluated using metrics such as PSNR, SSIM, and MSE, quantitatively measures the perceptual fidelity between a processed image and its reference by capturing how closely the enhanced or binarized output preserves the original visual appearance. PSNR and MSE assess pixel-level accuracy, while SSIM captures structural and perceptual similarity based on luminance and texture consistency [32]. Collectively, these metrics offer a comprehensive basis for assessing visual integrity and document readability.

MSE calculates the average of the squared differences between the reference image Y and the processed image Z, both of which are $m \times n$ in size. This is shown in Equation 5.

$$MSE = \frac{1}{N}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left(Y(i,j) - Z(i,j)\right)^2 \tag{5}$$

where $Y(i,j)$and $Z(i,j)$ are the pixel values at position (i, j) in images $Y$and Z, respectively. Here, $m$and $n$represent the numbers of rows and columns, and $N = m \times n$. In this study, Ycorresponds to the original image I_{ori}, while Z denotes the binarized image I_B [32].

As shown in Equation 7, PSNR measures how closely the binarized image I_B matches the original image I_{ori} [32].

$$PSNR = 10.\log_{10}\left(\frac{MAX^2}{MSE}\right) \tag{6}$$

where MAX is the maximum possible pixel value in the compared images.

SSIM measures the similarity of luminance, contrast, and structure, as shown in Equation 7.

$$SSIM(x,y) = \frac{\left(\left(2\,\mu_x\mu_y + c_1\right)\left(2\,\sigma_{xy} + c_2\right)\right)}{\left(\left(\mu_x^2 + \mu_y^2 + c_1\right)\left(\sigma_x^2 + \sigma_y^2 + c_2\right)\right)} \tag{7}$$

where μ_x and μ_y are the mean intensities, σ_x^2 and σ_y^2 are the variances, σ_{xy} is the covariance between x and y, and c_1 and c_2 are small constants introduced to stabilize the division [32].

Segmentation accuracy evaluation

In document image binarization, true positives (TP), false positives (FP), and false negatives (FN) are fundamental for computing Precision, Recall, and F1-Score, which measure the efficacy of text-background separation by reflecting the preservation of text strokes and the suppression of background noise; these metrics influence readability and OCR performance and offer an objective framework for evaluating binarization quality in automated document analysis. Distance Reciprocal Distortion (DRD) quantifies structural distortion in relation to binary ground truth by punishing misclassified pixels, particularly those adjacent to text boundaries; hence, a lower DRD signifies reduced distortions and enhanced preservation of character shapes [33], [34], [35].

In Equations 8 and 9, precision and recall are defined. TP stands for correctly detected foreground (text) pixels, FP stands for background pixels that were incorrectly classified as foreground, and FN stands for foreground pixels that were missed by the binarization [33].

$$\text{Precision} = \frac{TP}{TP + FP} \tag{8}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{9}$$

As shown in Equation 10, the F1-score combines Precision and Recall into one balanced measure [33].

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{10}$$

The Foreground Ratio (FG Ratio) indicates the percentage of pixels in the binarized result that are classified as foreground (text), as shown in Equation 11 [33].

$$FG_{ratio} = \frac{N_{FG}}{N_{total}} \times 100\% \tag{11}$$

Where N_{total} denotes the total number of pixels.

DRD measures structural distortion compared to the ground truth by punishing pixels that are incorrectly classified, especially those that affect character boundaries. A lower DRD means fewer distortions and better shape preservation, as shown in Equation 12.

$$DRD = \frac{(\sum_k DRD_k)}{NUBN} \tag{12}$$

where DRD_k is the distortion of each pixel that was misclassified, and $NUBN$ is the number of 8×8 blocks in the ground-truth image that are not entirely black or completely white [33].

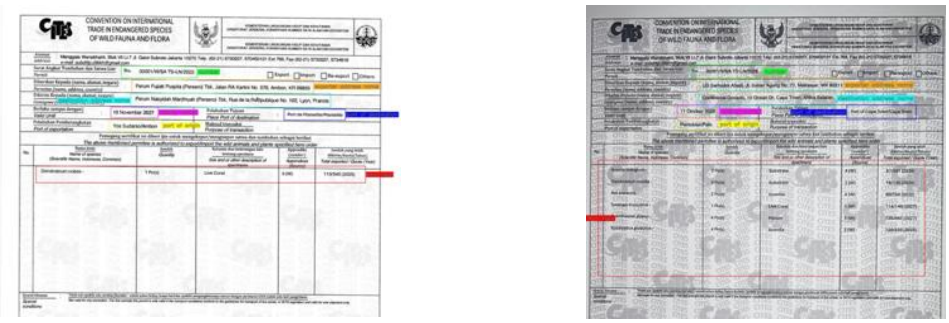
Implementation Consistency

Detectron2 inference consistently produces the same crops regardless of the conditions. All outputs and ground truth utilize the binary convention (0 = text/foreground, 255 = background), ensuring that pixel-based metrics can be compared. Otsu thresholds employ $\alpha \in \{0.9,1.0,1.1\}$, and Sauvola uses a 25x25 frame and $k \in \{0.3,0.4,0.5\}$. The clip limit for CLAHE is always 1.0, and the tile size is always 8x8.

RESULTS AND DISCUSSIONS

Detection and Cropping Image

Detectron2 detection generates bounding boxes that are spatially aligned with the fixed SATS-LN layout for the seven specified fields, namely *number, exporter_address_name, destination_address_name, validity_period, port_of_origin, port_of_destination,* and *item_table*. The resulting crops are uniform across both DOC and CAM domains, facilitating a controlled, field-level comparison of thresholding strategies under identical regional definitions. Figure 4 illustrates an example of the Detectron2 detection output, wherein each field is delineated by a labeled bounding box that specifies the cropping regions employed in subsequent binarization and evaluation procedures.



(a) DOC (b) CAM

Figure 4. Detectron2 bounding box detection for the seven SATS-LN fields

Binarization Dynamics Across Domains

To assess parameter sensitivity, each cropped field undergoes six binarization scenarios: three Otsu configurations (0.9×T_Otsu, 1.0×T_Otsu, 1.1×T_Otsu) and three Sauvola configurations (25×25 local window, k = 0.3, 0.4, and 0.5), followed by CLAHE (clip limit 1.0, tile size 8×8) for contrast normalization. Otsu remains stable in the DOC domain due to uniform lighting, but a slightly strict global threshold can erase tiny strokes. Sauvola better preserves local stroke structure in regions with mild shading and varying ink density. In the CAM domain, non-uniform illumination predominates, so a single global threshold cannot simultaneously handle glare and shadow, thereby reducing Otsu stability. Sauvola preserves stroke continuity across illumination by adapting thresholds to local statistics, and after CLAHE normalization, its outputs remain more consistent across crops

Figures 5 and 6 show the representative binarization results for all tested parameter values of Otsu and Sauvola, together with the related ground truth. They show the qualitative variations in stroke preservation and background suppression between DOC and CAM circumstances.

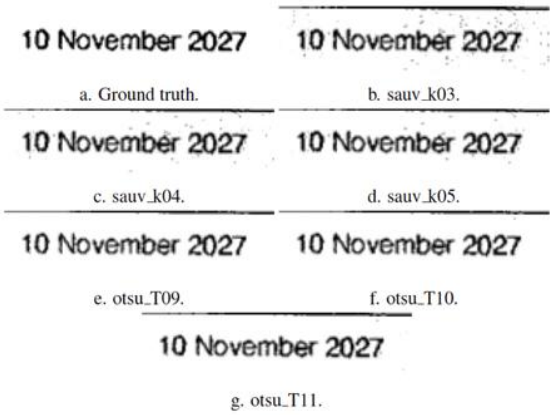


Figure 5. Otsu and Sauvola results with ground truth on DOC

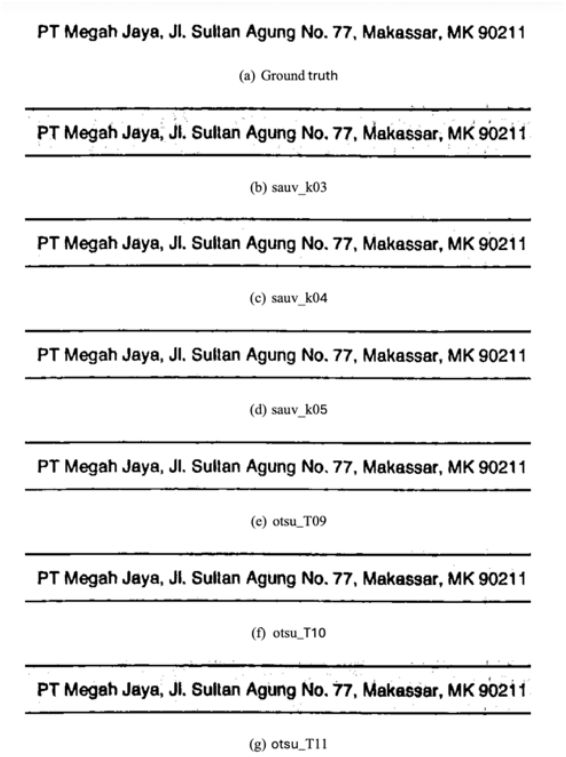


Figure 6. Otsu and Sauvola results with ground truth on CAM

According to prior research, global Otsu thresholding is more stable under relatively uniform lighting conditions, while adaptive Sauvola performs better under local illumination variations and non-uniform background shading. The generally consistent illumination in the DOC domain aids Otsu's global decision boundary, although strict thresholds may erode delicate character strokes. In the CAM domain, where shadows and glare are more prominent, Sauvola's local adaptation better maintains stroke continuity and eliminates irregular background artifacts, resulting in more consistent binarization across crops.

Visual Quality Analysis (PSNR, SSIM, and MSE)

Visual similarity assessment compares the binarized image with the original crop. Lower MSE and higher PSNR denote greater fidelity, whereas SSIM highlights the preservation of structural integrity. In the DOC domain, Otsu achieves a competitive PSNR due to its effective background removal under uniform illumination; however, Sauvola frequently produces higher SSIM values because it better preserves local structures. Within the CAM domain, Sauvola generally yields higher SSIM scores due to the adaptive thresholds' ability to maintain stroke continuity across illuminated and shadowed areas.

Table 1 indicates that global thresholding yields the highest PSNR and the lowest MSE on scanned images within the DOC domain under conditions of relatively uniform illumination. The highest median PSNR is achieved by Otsu_T10 (Median = 21.19 dB; IQR = 0.43), closely followed by Otsu_T09 (Median = 21.15 dB; IQR = 0.44). Additionally, the lowest MSE is also observed with Otsu_T10 (Median = 494.69; IQR = 49.62), indicating minimal intensity deviation from the original crop. However, structural similarity indicates that adaptive thresholding is preferable, with Sauvola_k03 achieving the highest SSIM (Median = 0.730; IQR = 0.048), whereas Otsu configurations remain comparatively lower (Median approximately 0.697–0.704). This suggests that Sauvola better preserves text structure, despite Otsu's superior performance in terms of PSNR and MSE.

Table 1. Visual metrics of DOC for otsu and sauvola

Domain	Method	PSNR		SSIM		MSE	
		Med	IQR	Med	IQR	Med	IQR
DOC	sauv_k03	19.07	0.48	0.730	0.048	806.07	89.21
	sauv_k04	19.70	0.43	0.719	0.055	696.91	69.26
	sauv_k05	20.14	0.41	0.712	0.060	629.99	59.82
	otsu_T09	21.15	0.44	0.697	0.066	498.80	50.59
	otsu_T10	21.19	0.43	0.700	0.066	494.69	49.62
	otsu_T11	21.03	0.42	0.704	0.063	513.02	49.36

Table 2 shows that the CAM domain displays greater illumination variability, as evidenced by consistently lower PSNR values across methods (Median approximately 11.77–11.84 dB) and comparatively higher MSE values. The highest median PSNR is achieved by Otsu_T10 (Median = 11.84 dB; IQR = 1.21). Nonetheless, structural similarity demonstrates greater stability under adaptive thresholding, with Sauvola_k03 attaining the highest SSIM (Median = 0.516; IQR = 0.037), surpassing the SSIM values observed with Otsu (Median approximately 0.506–0.510). Overall, these findings suggest that under CAM conditions, SSIM more accurately represents stroke preservation than PSNR alone, as global thresholds struggle to effectively address glint and shadow concurrently.

Table 2. Visual metrics of CAM for otsu and sauvola

Domain	Method	PSNR		SSIM		MSE	
		Med	IQR	Med	IQR	Med	IQR
CAM	sauv_k03	11.77	1.16	0.516	0.037	4324.38	1181.50
	sauv_k04	11.81	1.22	0.511	0.039	4290.90	1226.79
	sauv_k05	11.78	1.25	0.508	0.040	4315.75	1269.05
	otsu_T09	11.79	1.29	0.506	0.040	4311.02	1298.10
	otsu_T10	11.84	1.21	0.510	0.038	4252.43	1214.31
	otsu_T11	11.78	1.13	0.509	0.041	4315.50	1140.61

Segmentation Accuracy (Precision, Recall, F1, FG Ratio, and DRD)

Table 3 reveals that Otsu generally produces high precision but comparatively lower recall, suggesting that the background is effectively preserved at the expense of overlooking fine strokes. For instance, Otsu_T09 yields a Precision of 0.973 and a Recall of 0.868, resulting in an F1 score of 0.917. Conversely, Sauvola offers a more balanced compromise between precision and recall. The highest F1-score is attained by

Sauvola_k05 (Precision = 0.935; Recall = 0.941; F1 = 0.938), signifying a closer alignment with the ground truth. The FG Ratio for Sauvola (0.127–0.134) consistently exceeds that of Otsu (0.113–0.119), which aligns with enhanced preservation of stroke pixels in the binarized result. This trend is further corroborated by the DRD results, in which lower values signify fewer structural distortions relative to the ground truth. Among all DOC configurations, Sauvola_k05 produces the lowest DRD (Median = 1.52; IQR = 0.27), indicating the most precise preservation of character boundaries compared to other settings.

Table 3. Metrics for DOC segmentation using Otsu and Sauvola methods

Domain	Method	Performance Measurement			FG Ratio	DRD	
		Precision	Recall	F1		Med	IQR
DOC	sauv_k03	0.891	0.954	0.921	0.134	2.13	0.46
	sauv_k04	0.920	0.950	0.935	0.130	1.68	0.31
	sauv_k05	0.935	0.941	0.938	0.127	1.52	0.27
	otsu_T09	0.973	0.868	0.917	0.113	1.77	0.24
	otsu_T10	0.967	0.886	0.924	0.116	1.65	0.23
	otsu_T11	0.958	0.905	0.931	0.119	1.56	0.23

Table 4 illustrates the benefits of adaptive thresholding in the context of illumination irregularities within the CAM domain. Sauvola_k04 demonstrates superior segmentation performance, achieving a Precision of 0.972, a Recall of 0.989, and an F1 score of 0.980. Otsu_T11 achieves a high recall of 0.994; however, its precision significantly decreases to 0.860, leading to a reduced F1-score of 0.920. This indicates potential over-segmentation, likely due to illumination artifacts. The FG Ratio reinforces this analysis: Otsu_T11 presents the highest FG Ratio (0.175), indicative of significant foreground activation, while Sauvola_k04 exhibits a more balanced FG Ratio (0.156). The DRD values provide additional insight, with lower distortion signifying a closer alignment to the ground truth. Sauvola_k04 exhibits the lowest DRD (Med = 0.53; IQR = 0.11), whereas Otsu_T11 presents the highest DRD (Med = 2.03; IQR = 0.71), thereby confirming a higher degree of structural distortion associated with strict global thresholding in CAM conditions.

Table 4. Metrics for CAM segmentation using otsu and sauvola methods

Domain	Method	Performance Measurement			FG Ratio	DRD	
		Precision	Recall	F1		Med	IQR
CAM	sauv_k03	0.910	0.992	0.949	0.166	1.41	0.40
	sauv_k04	0.972	0.989	0.980	0.156	0.53	0.11
	sauv_k05	0.988	0.947	0.967	0.146	0.82	0.17
	otsu_T09	0.973	0.925	0.948	0.145	1.23	0.30
	otsu_T10	0.934	0.975	0.953	0.160	1.13	0.37
	otsu_T11	0.860	0.994	0.920	0.175	2.03	0.71

Figure 7 shows a comparison of the F1-score performance of different binarization methods for the two acquisition domains. This is done to see how well each method keeps the separation between text and background.

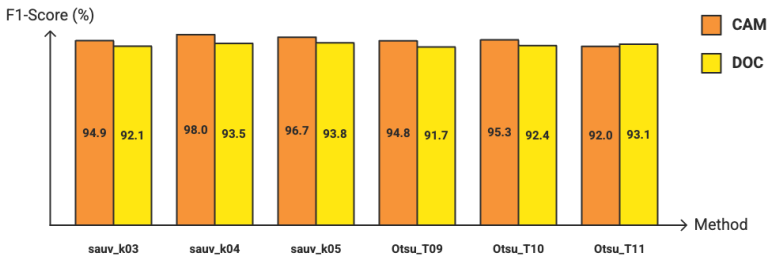


Figure 7. Comparison of F1-Score accuracy based on domain and method

Overall, the results show that the F1-scores in the CAM domain are always higher than those in the DOC domain. This means that adaptive thresholding works better at keeping text safe when the lighting changes. In the DOC domain, Sauvola_k05 achieves the highest F1-score (F1 = 0.938), reflecting a balanced trade-off between precision and recall, while Otsu-based methods exhibit slightly lower performance due to their tendency to miss fine strokes. On the other hand, in the CAM domain, Sauvola_k04 gets the best F1-score

(F1 = 0.980). This shows that local adaptive thresholding works better than strict global thresholds for dealing with uneven lighting and cutting down on over-segmentation.

For further research, it is recommended to examine how the quality of the detection stage cropping (e.g., localization accuracy and bounding-box tightness) influences binarization outcomes, since cropped regions serve as the input for the thresholding process. In addition, future studies should validate the proposed preprocessing guideline on a wider range of document types with different layouts and degradation characteristics to assess its generalizability beyond SATS-LN. These extensions are expected to strengthen the applicability of the domain-aware binarization strategy for large-scale structured document digitization under heterogeneous acquisition conditions.

CONCLUSION

This study assessed the impact of a crop-first SATS-LN preprocessing pipeline Detectron2-based element detection followed by Otsu/Sauvola binarization and CLAHE using two complementary evaluation models (visual similarity and segmentation accuracy) on 200 documents (150 DOC scans and 50 CAM photos), and the results confirm that method preference depends on the evaluation objective and acquisition domain: in DOC images, Otsu can appear advantageous in visual fidelity (e.g., Otsu_T10 achieving the highest median PSNR of 21.19 dB with the lowest median MSE of 494.69), yet segmentation-based evidence shows that Sauvola better preserves character strokes and conforms more closely to binary ground truth, with Sauvola_k05 producing the best DOC separation performance (F1 = 0.938), while in CAM images with strong illumination variability, Sauvola's superiority is more apparent across structural and segmentation indicators, where Sauvola_k04 yields the strongest overall performance (F1 = 0.980) and avoids the over-segmentation tendency observed in strict global thresholds (e.g., Otsu_T11 precision dropping to 0.860 despite recall of 0.994). The experimental configuration, employing a 25x25 Sauvola window, a 1.0 CLAHE clip limit, and an 8x8 tile size, indicates that Sauvola_k05 and Sauvola_k04 are optimal for DOC and CAM, respectively, in terms of preserving textual fidelity and minimizing background artifacts. An evidence-based binarization guideline, designed for deployment in the context of structured government document digitization, enhances text background separation across diverse capture conditions and augments subsequent automated document analysis systems that leverage clean, stroke-preserving binary representations. This guideline delivers practical impact by minimizing manual rework, improving cross-device preprocessing consistency, and enabling more reliable large-scale extraction and archiving of structured government documents. For future research, this framework can be extended to other structured document types and combined with adaptive or learning-based threshold selection strategies to further enhance robustness across broader degradation conditions.

REFERENCES

[1] O. Savchenko, “Innovative aspects of development of digitalization of public governance in the USA,” *Democratic governance*, vol. 30, no. 2, pp. 120–130, 2022, doi: 10.23939/dg2022.02.120.

[2] S. Lafia, D. A. Bleckley, and J. T. Alexander, “Digitizing and parsing semi-structured historical administrative documents from the G.I. Bill mortgage guarantee program,” *Journal of Documentation*, vol. 79, no. 7, pp. 225–239, 2023, doi: 10.1108/JD-03-2023-0055.

[3] Y. Naik and Yogish Naik G. R., “Enhancement of Degraded Historical Document Images for Binarization,” *Journal of Electrical Systems*, vol. 20, no. 3, pp. 4779–4796, 2024, doi: 10.52783/jes.5990.

[4] S. Guan *et al.*, “PreP-OCR: A Complete Pipeline for Document Image Restoration and Enhanced OCR Accuracy,” 2025. doi: 10.18653/v1/2025.acl-long.749.

[5] S. Dey and P. Jawanpuria, “Confidence Score for Unsupervised Foreground Background Separation of Document Images,” 2022. doi: 10.48550/arxiv.2204.04044.

[6] D. Asatryan, M. Haroutunian, G. Sazhumyan, A. Kupriyanov, R. Paringer, and D. Kirsh, “Improving Binarization Methods for Historical Handwritten Documents,” in *Proceedings of Computer Science and Information Technologies 2023 Conference*, 2023. doi: 10.51408/csit2023_54.

[7] C. Wibawa and D. T. Anggraeni, “Comparison of Image Segmentation Method in Image Character Extraction Preprocessing Using Optical Character Recoginiton,” *Jurnal Teknik Informatika (Jutif)*, vol. 4, no. 3, pp. 583–589, 2023, doi: 10.52436/1.jutif.2023.4.3.956.

[8] B. Bataineh *et al.*, "A Comprehensive Review on Document Image Binarization," *J. Imaging*, vol. 11, no. 5, 2025, doi: 10.3390/jimaging11050133.

[9] D. Li, Y. Wu, and Y. Zhou, "SauvolaNet: Learning Adaptive Sauvola Network for Degraded Document Binarization," 2021. doi: 10.1007/978-3-030-86337-1_36.

[10] Y. Tian and M. Han, "Adaptive Binarization for Vehicle State Images Based on Contrast Preserving Decolorization and Major Cluster Estimation," *IEICE Trans. Inf. Syst.*, vol. E105D, no. 3, pp. 679–688, 2022, doi: 10.1587/transinf.2021EDP7218.

[11] G. R. Mukarambi, "Hybrid Method for Elimination of Uneven Illumination from Camera-based Document Images," *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 3, pp. 3566–3570, Jan. 2020, doi: 10.35940/ijitee.B7517.019320.

[12] I. Mohammed, N. Isa, I. Majid Mohammed, and N. Ashidi Mat Isa, "Contrast Limited Adaptive Local Histogram Equalization Method for Poor Contrast Image Enhancement," *IEEE Access*, vol. 13, pp. 62600–62632, 2025, doi: 10.1109/ACCESS.2025.3558506.

[13] Z. Jia, Z. Shi, Z. Quan, and M. Shunqi, "Fabric defect detection based on transfer learning and improved Faster R-CNN," *J. Eng. Fiber. Fabr.*, vol. 17, 2022, doi: 10.1177/15589250221086647.

[14] B. Bataineh *et al.*, "Efficient Text Bounding Box Identification Using Mask R-CNN: Case of Thai Documents," *IEEE Access*, vol. 12, no. 1, pp. 40701–40743, Nov. 2023, doi: 10.1109/ACCESS.2024.3383911.

[15] S. Naik, K. A. Hashmi, A. Pagani, M. Liwicki, D. Stricker, and M. Z. Afzal, "Investigating Attention Mechanism for Page Object Detection in Document Images," *Applied Sciences (Switzerland)*, vol. 12, no. 15, p. 7486, Jul. 2022, doi: 10.3390/app12157486.

[16] M. Zhang, Y. Su, and X. Hu, "Small target detection based on faster R-CNN," 2023, p. 127. doi: 10.1117/12.2660388.

[17] X. Ding, Q. Li, Y. Cheng, J. Wang, W. Bian, and B. Jie, "Local keypoint-based Faster R-CNN," *Applied Intelligence*, vol. 50, no. 10, pp. 3007–3022, 2020, doi: 10.1007/s10489-020-01665-9.

[18] L. Yabin, Y. Jun, and H. Zhiyi, "Improved Faster R-CNN Algorithm for Sea Object Detection Under Complex Sea Conditions," *International Journal of Advanced Network, Monitoring and Controls*, vol. 5, no. 2, pp. 76–82, 2020, doi: 10.21307/ijanmc-2020-020.

[19] Y. Zhang and T. Lu, "RecFRCN: Few-Shot Object Detection With Recalibrated Faster R-CNN," *IEEE Access*, vol. 11, pp. 121109–121117, 2023, doi: 10.1109/ACCESS.2023.3328390.

[20] J. Liu, "Face recognition technology based on ResNet-50," *Applied and Computational Engineering*, vol. 39, no. 1, pp. 160–165, 2024, doi: 10.54254/2755-2721/39/20230593.

[21] D. Guo, Z. Wu, J. Feng, and T. Zou, "Multi-scale semantic enhancement network for object detection," *Sci. Rep.*, vol. 13, no. 1, 2023, doi: 10.1038/s41598-023-34277-7.

[22] E. I. Ershov, S. A. Korchagin, V. V. Kokhan, and P. V. Bezmaternykh, "A generalization of otsu method for linear separation of two unbalanced classes in document image binarization," *Computer Optics*, vol. 45, no. 1, pp. 66–76, 2021, doi: 10.18287/2412-6179-CO-752.

[23] X. Liu, Z. Zhang, Y. Hao, H. Zhao, and Y. Yang, "Optimized OTSU Segmentation Algorithm-Based Temperature Feature Extraction Method for Infrared Images of Electrical Equipment," *Sensors*, vol. 24, no. 4, 2024, doi: 10.3390/s24041126.

[24] W. A. H. Jumiawi and A. El-Zaart, "Improvement in the Between-Class Variance Based on Lognormal Distribution for Accurate Image Segmentation," *Entropy*, vol. 24, no. 9, 2022, doi: 10.3390/e24091204.

[25] Z. Y. Tan, S. N. Basah, H. Yazid, and M. J. A. Safar, "Performance analysis of Otsu thresholding for sign language segmentation," *Multimed. Tools Appl.*, vol. 80, no. 14, pp. 21499–21520, 2021, doi: 10.1007/s11042-021-10688-4.

[26] M. Chandrakala, "Image Analysis of Sauvola and Niblack Thresholding Techniques," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 9, no. VI, pp. 2353–2357, 2021, doi: 10.22214/ijraset.2021.34569.

[27] F. F. Alkhalid, A. M. Hasan, and A. A. Alhamady, "Improving radiographic image contrast using multi layers of histogram equalization technique," *IAES International Journal of Artificial Intelligence*, vol. 10, no. 1, pp. 151–156, 2021, doi: 10.11591/ijai.v10.i1.pp151-156.

[28] O. Kamel, K. Amin, N. Semary, and N. Aboelenien, "An Automated Contrast Enhancement Technique for Remote Sensed Images," *IJCI. International Journal of Computers and Information*, vol. 0, no. 0, pp. 0–0, 2023, doi: 10.21608/ijci.2023.212239.1111.

[29] T. C. Tung and C. S. Fuh, "ICEBIN: Image Contrast Enhancement Based on Induced Norm and Local Patch Approaches," *IEEE Access*, vol. 9, pp. 23737–23750, 2021, doi: 10.1109/ACCESS.2021.3056244.

[30] K. Nikolaidou, M. Seuret, H. Mokayed, and M. Liwicki, "A survey of historical document image datasets," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 25, no. 4, pp. 305–338, Dec. 2022, doi: 10.1007/s10032-022-00405-8.

[31] C. Tensmeyer and T. Martinez, "Historical Document Image Binarization: A Review," *SN Comput. Sci.*, vol. 1, no. 3, p. 173, May 2020, doi: 10.1007/s42979-020-00176-1.

[32] A. A. Cardona-Mesa, R. D. Vásquez-Salazar, J. P. Diaz-Paz, H. O. Sarmiento-Maldonado, L. Gómez, and C. M. Travieso-González, "Optimization of Autoencoders for Speckle Reduction in SAR Imagery Through Variance Analysis and Quantitative Evaluation," *Mathematics*, vol. 13, no. 3, pp. 1–27, 2025, doi: 10.3390/math13030457.

[33] A. Sulaiman, K. Omar, and M. F. Nasrudin, "Degraded Historical Document Binarization: A Review on Issues, Challenges, Techniques, and Future Directions," *J. Imaging*, vol. 5, no. 4, p. 48, Apr. 2019, doi: 10.3390/jimaging5040048.

[34] R. Paudyal and D. Pyakurel, "Enhancing the Efficiency of Deep Learning Models for Handwritten Text Recognition by Utilizing Meta-learning Optimization Techniques," *Journal of Advanced College of Engineering and Management*, vol. 9, pp. 1–13, Nov. 2024, doi: 10.3126/jacem.v9i1.71399.

[35] Y. Zhou, S. Zuo, Z. Yang, J. He, J. Shi, and R. Zhang, "A Review of Document Image Enhancement Based on Document Degradation Problem," *Applied Sciences (Switzerland)*, vol. 13, no. 13, 2023, doi: 10.3390/app13137855.