

Review

A Review of Document Binarization: Main Techniques, New Challenges, and Trends

Zhengxian Yang, Shikai Zuo *, Yanxi Zhou, Jinlong He and Jianwen Shi

School of Opto-Electronic and Communication Engineering, Department of Microelectronics, Xiamen University of Technology, Xiamen 361024, China; 2222031314@s.xmut.edu.cn (Z.Y.); 2222031329@s.xmut.edu.cn (Y.Z.); 2222031247@s.xmut.edu.cn (J.H.); 2222031285@s.xmut.edu.cn (J.S.)

* Correspondence: skzuo@xmut.edu.cn; Tel.: +86-159-8088-8371

Abstract: Document image binarization is a challenging task, especially when it comes to text segmentation in degraded document images. The binarization, as a pre-processing step of Optical Character Recognition (OCR), is one of the most fundamental and commonly used segmentation methods. It separates the foreground text from the background of the document image to facilitate subsequent image processing. In view of the different degradation degrees of document images, researchers have proposed a variety of solutions. In this paper, we have summarized some challenges and difficulties in the field of document image binarization. Approximately 60 methods documenting image binarization techniques are mentioned, including traditional algorithms and deep learning-based algorithms. Here, we evaluated the performance of 25 image binarization techniques on the H-DIBCO2016 dataset to provide some help for future research.

Keywords: degraded document images; binarization; threshold processing; deep learning



Citation: Yang, Z.; Zuo, S.; Zhou, Y.; He, J.; Shi, J. A Review of Document Binarization: Main Techniques, New Challenges, and Trends. *Electronics* **2024**, *13*, 1394. <https://doi.org/10.3390/electronics13071394>

Academic Editor: Stefanos Kollias

Received: 18 January 2024

Revised: 9 March 2024

Accepted: 27 March 2024

Published: 7 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Image binarization is an important aspect of image analysis, such as scene text detection [1–3] and medical image analysis [4,5]. Especially in the field of document image processing, binarization has a wide range of applications as a basic method of digital image processing, including text recognition, document image segmentation, image morphological processing, and feature extraction [6–9]. It commonly serves as the primary stage in document analysis and recognition systems, as well as Optical Character Recognition (OCR), exerting a substantial impact on the efficacy of subsequent character segmentation and recognition.

Binarization is a technique used to separate the region of interest, such as text, from the background in an image, and it represents one of the fundamental methods of image segmentation. This process involves converting a grayscale image into a binary black and white image. Thresholding is a widely used tool in image segmentation [10], including global, local, and various automatic thresholding methods [11,12]. However, these approaches encountered difficulties in effectively handling complex images with uneven pixel distributions and noise interference. In recent years, the advancement of deep learning technology has led to the successful application of deep learning algorithms, including Convolutional Neural Networks (CNNs) [13–16], Generative Adversarial Networks (GANs) [17,18], and U-Net [19] in the field of document image binarization. The continuous emergence of new algorithms and technologies offers opportunities for further optimization in the field of document image binarization. Despite extensive research in digital image processing, there are still numerous unresolved challenges in handling degraded document images.

The main difficulties encountered in document image binarization are related to the non-uniform variations present in the image, as illustrated in Figure 1. Particularly in degraded document images, a range of issues such as aging, damage, blurring, and fading are frequently encountered. These challenges not only diminish the overall image

quality but also render binarization algorithms more complex in handling these irregular alterations. Moreover, low-quality document images often exhibit various imperfections, including aged paper and documents, noise introduced during the scanning process (such as Gaussian noise, white noise, salt-and-pepper noise), ink stains, and contamination. These issues lead to the emergence of numerous isolated points and abnormal areas during the binarization process. Additionally, document images are susceptible to variations in lighting and contrast, resulting in uneven brightness and color distribution, thereby amplifying the complexity of binarization. Common problems such as spot defects and fractures also give rise to disconnected areas in the binarized output, making it challenging to accurately extract the image content. Displacement, skew, and deformation are prevalent issues during the document scanning process. These deformations can cause distortion in the document image during binarization, directly impacting subsequent document processing and analysis.



Figure 1. Original images from (H)DIBCO datasets [20–24].

Previous reviews on document image binarization have been much narrower in scope compared to the current article. Some reviews [25–27] choose to delve into detailed discussions and comparisons of various binarization methods, aiming to provide readers with concise choices. Some reviews [28–30] focus on threshold-based binarization techniques, overlooking popular neural network algorithms. Therefore, we aim to provide a broader and more comprehensive review of binarization techniques, offering researchers an extensive overview of binarization technologies. This paper seeks to conduct a comprehensive review and synthesis of prevalent methods for document image binarization within an open research framework. The review encompasses both conventional algorithms and deep learning-based approaches, with the objective of furnishing valuable insights for prospective investigations in the field of document image binarization.

2. Traditional Binarization Techniques

The threshold method is an image segmentation technique that relies on the grayscale value of a pixel to separate the image based on a specified threshold. This method typically involves two technical approaches: the global thresholding method and the local thresholding method, which separate the image pixels based on their size relationship to the specified threshold.

2.1. Global Threshold Method

The Otsu [31] algorithm, developed in 1979, is a prominent method of a global thresholding technique. The algorithm aims to determine an optimal threshold value, denoted as T , by analyzing the grayscale properties of an image. This process involves partitioning the image into foreground and background segments. The objective is to minimize the gap between the two segments while maximizing the difference between them. The difference in grayscale distribution serves as a measure of the contrast between foreground and background, with a larger difference indicating an easier segmentation. The Otsu algorithm is also commonly known as the maximized difference between classes method. The optimal threshold for the desired image is the value that maximizes the gap between categories, and it can be expressed as follows:

$$T' = \arg \max_{0 \leq T \leq L} \omega_0(T)\omega_1(T)(\mu_0(T) - \mu_1(T))^2, \quad (1)$$

we represent the image pixel in the gray level of the image, the image has L -order gray level, $\omega_0(T)$ and $\omega_1(T)$ are the probability distribution of the target and background when the threshold value is T , $\mu_0(T)$ and $\mu_1(T)$ represent the average gray value of the pixel of the target and background, respectively, if the pixel value of the input image is greater than T' . The pixel value is set to white, or otherwise it is black.

The Otsu algorithm partitions the entire image based on a single threshold, allowing for the determination of the optimal threshold for the image at once. This approach generally yields improved separation for images with a uniform background. However, it may result in suboptimal image processing for images with uneven backgrounds, such as misidentification of the background in document images with significant ink penetration or insufficient grayscale contrast. Figure 2 shows a few examples of the Otsu algorithm applied on document images.

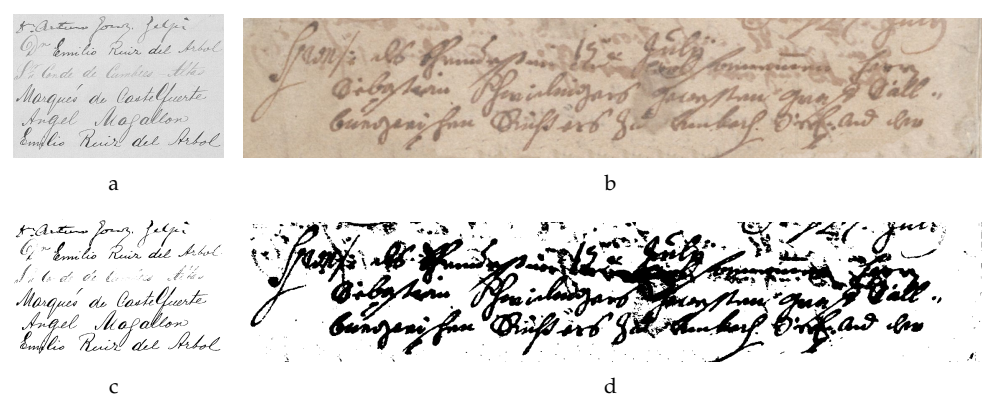


Figure 2. Original images from HDIBCO2016 dataset [23] and binarization results of Otsu's. (a) Original image; (b) Original image; (c) binarization result of (a); (d) binarization result of (b).

As demonstrated by the most representative global method, Otsu, such approaches are unsuitable for low-contrast or uneven images. However, using it on documents with a uniformly pure background may indeed be a good choice.

2.2. Local Threshold Method

The Niblack [32] algorithm was developed to address the limitations of a fixed threshold by introducing a local binarization method. This approach involves utilizing a local window to calculate the mean and standard deviation within a small neighboring domain of each pixel. These values are then used to adjust the threshold for binarizing the image. The threshold calculation formula is expressed as follows:

$$T = m + k \times s, \quad (2)$$

where m represents the average gray value of pixels in the local area, s represents the standard deviation, and k is a constant correction factor, which can be adjusted according to the foreground and background conditions of the image. Trier [33] believes that Niblack performs better than other local binarization methods in gray images with low contrast, noise, and uneven background intensity.

The Sauvola [34] algorithm is an improvement upon the Niblack algorithm, designed to address the problem of excessive noise levels. It introduces a new parameter, R , which is based on the dynamic range of the standard deviation. The threshold calculation formula is expressed as follows:

$$T = m \times \left[1 + k \times \left(\frac{s}{R} - 1 \right) \right], \quad (3)$$

it can be seen from the formula that Sauvola introduces a new parameter R , which represents the dynamic range of standard deviation. In regions with high contrast in the pixel neighborhood, when s approaches R , the threshold value approximately equals the average value. This makes it unfriendly to high-contrast regions. Therefore, it is necessary to choose the optimal values based on the characteristics and distribution of the image. However, this requires manually determining the values of the factor k and the window size. Additionally, it faces challenges in handling targets of different sizes and accurately capturing all characters when different font sizes are present in the same text [35].

In addressing the issue of black noise in the Niblack algorithm, Khurshid et al. [36] introduced an algorithm named “NICK”, which is purported to be more effective for deteriorated and noisy antique documents. In comparison to Niblack, it offers the advantage of significantly improving the binarization of light-colored page region images by reducing the binary threshold. The formula for calculating the threshold is expressed as follows:

$$T = m + k \sqrt{\frac{\sum p_i^2 - m^2}{NP}}, \quad (4)$$

p_i represents the pixel value of a grayscale image, while NP denotes the number of pixels. The presence of noise can be reduced when the k value approaches 0.2, but this may lead to interrupted or faint characters. When the k value is close to 0.1, the text can be extracted but some noise is retained. It also did not address the issue of manually determining the factor. Consequently, B. Bataineh [37] contends that the method does not outperform the Niblack algorithm in exceptional circumstances, such as very low-contrast images or the variations of text size and thickness.

Saxena [30] believes that the window size is the main defect of the local threshold method. Both large and small size windows will generate noise. Small windows are effective in removing noise but may distort the text, while large windows can effectively preserve the text but may introduce some noise. And even in windows without target pixels, the local threshold method can still detect target pixels. Bataineh et al. [37] proposed a threshold approach based on dynamic flexible windows. This method involves two approaches: dynamically segmenting images into windows based on image characteristics and determining the appropriate threshold for each window. The window size generally also affects the computation time. However, T. Romen Singh et al. [38] chose to utilize the dotted image as an initial stage in the calculation of the local mean. This approach allows

for the calculation of the average value to be independent of the window size. Compared to other local threshold techniques, this method does not involve the computation of standard deviations, thereby reducing computational complexity and accelerating processing speed.

Chaki [39] suggests that a larger value of k adds more pixels to the document image, thereby reducing text readability. Conversely, a smaller k value results in missing or incomplete characters, which reduces the number of potential pixels. Consequently, determining the appropriate value of k also becomes challenging. Taking the Niblack method as an example, Figure 3 illustrates the impact of different window sizes and k values on the local threshold method. Even with an accurate k value, it still generates pepper noise in the shadowed areas of the image or in non-text regions.

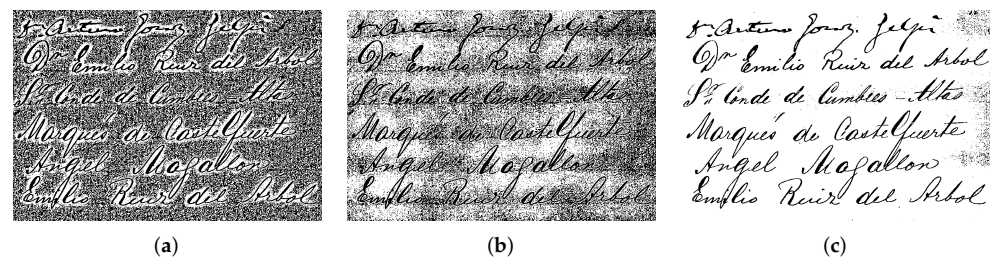


Figure 3. Niblack's binarization results for different window sizes and k values. (a) windows = 25×25 , $k = 0.2$; (b) windows = 125×125 , $k = 0.2$; (c) windows = 125×125 , $k = -0.8$.

Researchers have applied some adaptive improvements that enable these parameters to be automatically adjusted without human intervention. He [40] compared Niblack, Sauvola, and their adaptive threshold method in an article, and found that adaptive Niblack and adaptive Sauvola performed better than originals. In the paper [41], the authors combine the image contrast defined by local image minimum and maximum values with the computed Sauvola's binarization step, without manually adjusting the user-defined parameters to the document content. Figure 4 shows an example of this method. Usually, the fixed threshold is manually set according to the specific situation of different tasks, whereas the adaptive threshold tends to estimate the background surface of the document first, and then the thresholds are calculated according to the estimated background surface. For example, Lu et al. [42] incorporates the step of estimating the document background surface. Krzysztof et al. [43] also perform fingerprint background estimation as a preliminary step in the binarization task. They all utilize polynomial smoothing to estimate the background surface. This method is a direct background subtraction approach that does not require any rough estimates of foreground and background regions. Moghaddam et al. [44] estimate the backdrop surface of the document by an adaptive and iterative image averaging approach. Gatos [45] roughly divides images into background and foreground pixels and estimates a background surface by interpolating neighboring background intensities. Creating a local threshold policy involves combining the calculated background surface with the original image and integrating image up-sampling techniques. However, in He's study [40], it was found that removing the background from images before applying the Niblack and Sauvola algorithms did not bring any benefit in better binarization. This is because the two algorithms rely on certain information hidden in the background for local threshold calculation. Therefore, flexibility and adaptability should be exercised when applying them in practical applications.

In order to address the limitations of the global threshold method in dealing with complex structured documents, the local threshold method assigns separate thresholds to each window of the image. This method has low complexity and generally performs well in most cases. However, a drawback is the poor connectivity of segmented objects, which often results in the phenomenon of ghosting such as background areas exhibiting pseudo-strokes, leading to significant sensitivity to noise. Many studies first introduce additional techniques to handle such issues before utilizing global or local thresholding

methods, as seen in Su’s approach [46,47]. By relying on a combination of other techniques, complex methods can address degradation in document images, but they often require more computational resources. Here, we summarize some common threshold methods in Table 1.

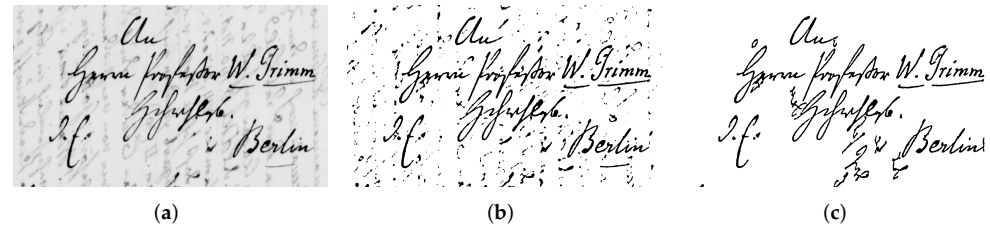


Figure 4. Examples of results from adaptive methods on HDIBCO2016 dataset. (a) original image; (b) Sauvola’s [34]; (c) improved Sauvola’s [41].

Table 1. Traditional binarization techniques (1).

Classification	Algorithm	Description	Performance
Global Threshold	Otsu [31]	The gray level corresponding to the maximum inter-class variance is selected as the global threshold.	Low complexity and fast operation. However, it cannot handle complex degraded images and is suitable for processing high-quality document images.
	Niblack [32]	It calculates the mean and standard deviation of the pixel within a local window of an image, laying the foundation for local binarization methods.	The processing time has increased, and obvious noise can be seen in the output binarization image, which greatly increases the foreground region.
	Bernsen [48]	Compute a separate threshold for each pixel by estimating and utilizing the local contrast before classification.	When dealing with complex backgrounds in document images, pseudo-shadow artifacts may occur.
	Sauvola [34]	It improves the Niblack method by modifying the threshold calculation.	High complexity. Performance degrades in high contrast regions, but it can eliminate dark areas in the background.
	Wolf [49]	Global statistical normalization based on Sauvola for automatic detection of text regions.	It can mitigate the impact of background noise but degradation in performance occurs when there is a sharp change in background gray values across the image.
Local Threshold	Gatos [45]	Local adaptive-based approach to enhance degraded image binarization documents.	It performs well on degraded document images with issues such as shadows, uneven lighting, and smear, but leaves a slight amount of noise.
	NICK [36]	In order to solve the black noise problem in Niblack, an improved method of threshold calculations was used.	It lowers the threshold and performs better in low-intensity images.
	Su [46]	Image contrast technique defined using local image maxima and minima.	This method can handle general document images but performs poorly in specific cases such as the ink-bleeding.
	Bataineh [37]	Dynamic segmentation of the image into windows based on image features, determining the threshold value for each window.	It can address specific challenges, such as thin pen strokes and low-contrast images. However, it unavoidably retains excess background.
	T.R.Singh [38]	Local adaptive threshold segmentation uses the local mean and mean difference to remove the background.	Unable to recognize faint text, it loses some fine details but exhibits a certain degree of recovery for unevenly lit backgrounds.
	Su [47]	Combination method based on local image contrast and local image gradient.	Compared to several classical thresholding methods, it exhibits superior visual quality. However, there is still a small amount of noise present.
	Hadjadj [41]	An improved Sauvola’s Algorithm technique for document images is presented.	It improves the quality of binarization results without adjusting manually the user-defined parameters to the document content.
	WAN [26]	Improving lost detail strokes by increasing the binarization threshold based on Sauvola.	Performs well on a clean background without pollution.

2.3. Mixed Threshold Method

To make up for the limitations of global and local thresholding approaches, researchers have proposed hybrid thresholding binarization algorithms. For example, Yang et al. [50] integrated Otsu and Bernsen's method. Zemouri et al. [51] enhanced the document picture using global thresholding before binarization and then applied a local thresholding strategy for binarization. Chaudhary et al. [52] developed a rudimentary estimation of the backdrop, constructed an image with a high contrast, and then thresholded it using the hybrid technique. Due to the low identification rate of blurred letters in handwritten document images, K. Ntirogiannis et al. [53] devised a blend of global and local adaptive binarization. First, a background estimate with picture normalization based on background compensation is applied. Then, global binarization is performed on the normalized image. In the binarized picture, typical attributes of the document image such as stroke width and contrast are determined. In addition, local adaptive binarization is performed on the normalized image. Finally, the results of the two binarizations are mixed. Liang [54] developed a hybrid thresholding technique and determined the trade-off between local and global content using variational optimization. Xiao et al. [55] suggested a model consisting of a global branch and a local branch that takes the global block of the downsampled picture and the local block of the source image as inputs correspondingly. The ultimate binarization is achieved by merging the findings of these two branches. Saddami et al. [56] employed an integrated technique such as local and global thresholding methods to extract text from the backdrop to recover the information on degraded ancient Jawi manuscripts. P. Ranjitha et al. [57] suggested a classification system to deal with degraded document photographs by blending the modification of local and global binarization algorithms.

It is easy to observe that these methods, which combine global and local thresholds, first use a global threshold to process the entire document and then use local thresholds to address the shortcomings of the global method. We summarize some mixed methods in Table 2.

Table 2. Traditional binarization techniques (2).

Classification	Algorithm	Description	Performance
Mixed threshold	Yang [50]	A combined Otsu and Bernsen method for processing printed document images.	It can repair broken strokes and eliminate minor ghost artifacts, but with increased processing time.
	Zemouri [51]	Apply the global threshold to the whole document, and use the Sauvola method on the intermediate image.	Its recognition effect is better than that of a single threshold, but the effect of degraded document images is not good.
	Chaudhary [52]	Otsu's method separates background pixels from the foreground. Sauvola's method is then applied to eliminate pixels erroneously estimated as foreground.	Thin or weak strokes can be easily confused with the background, making them difficult to identify. Suitable for high-contrast document images.
	K.Ntirogiannis [53]	Combining the results of both Otsu and Niblack, incorporating post-processing in intermediate and final steps.	It cannot simultaneously consider text information and large noise, making it unsuitable for document images with folds or page splits.
	Liang [54]	Combining the Otsu and Sauvola methods to determine thresholds.	It can reach the general level in document image binarization.
	Saddami [56]	It combines the Niblack and Wolf methods. The Otsu method and standard deviation are combined.	It retains a significant amount of noise and degrades the performance in images with low contrast.

2.4. Image Feature Method

Document images contain a wealth of information and are highly complex. Especially for degraded images, relying solely on thresholding is often insufficient. Therefore, pre-

processing or post-processing steps are commonly incorporated during binarization. In the work of Moghaddam et al. [44], these methods are categorized into four meta-levels: pixel-level, region-level, content-level, and global-level. Pixel-level characteristics involve factors such as grayscale values and gradients. Region-level is applicable to various image processing domains. Some basic examples include calculations such as mean and variance. Content-level, more suitable for document images, focuses on stroke information, including stroke grayscale and contours. The highest level is the global level, encompassing all data in the image.

In degraded document images, attention is paid to stroke continuity, especially in handwritten documents. Edge detection is a pixel-level method used for this purpose. Based on the threshold method of edge detection, it first identifies the edge pixels within the image and then uses these edges as partition boundaries to divide the image into different areas. Edge detection typically involves using differential operators to identify areas of significant variation in the grayscale values of images. The example of edge detection is shown in Figure 5. Commonly used edge detectors include Sobel, Prewitt, Roberts, Laplace of Gaussian, and Canny. The selection of the edge detector is determined by the specific characteristics of the image in practical applications. Santhanaprabhu et al. [58] applied the Sobel edge detection technique to extract text and perform document image binarization. Lu et al. [42] used the L1-norm image gradient to identify the edge of the font from the compensated document image. T. Lelore et al. [59] have described a quick solution for repairing document images by employing the edge-based method to locate text in degraded document images. However, it should be noted that the edges detected by the edge detection technique may not completely enclose the prospective text area, thus requiring further improvement. Holambe et al. [60] exploited adaptive image contrast in combination with Canny's edge diagram to identify the edge pixels of the font. Jia et al. [61] used structural symmetry pixels (SSPs) to calculate local thresholds for the neighborhood. SSPs is defined as the pixel around the stroke, whose gradient size is large enough and the direction is symmetric and opposite. The author extracts SSPs by combining the adaptive gradient binarization and iterative stroke width estimation algorithm. This approach reduces the influence of degraded documents and ensures the appropriate field size when determining the direction. Multi-threshold voting is then used to determine whether the pixel belongs to the foreground text, handling inaccurate SSPs detection. Hadjadj et al. [62] introduced a method of document image binarization that is frequently applied to the active contour model used in image segmentation. The objective of their method is widely used in the field of image segmentation. It aims to convert the problem of image segmentation to solve the minimum energy functional. Hadjadj defines image contrast of the maximum and minimum values of the local image. Use it to automatically generate initialization graphics for active contour models. The average threshold value is selected to generate binarization, as it enables the active contour to effectively detect low-contrast regions. When the active contour remains stationary, the result is obtained by thresholding the level set function.

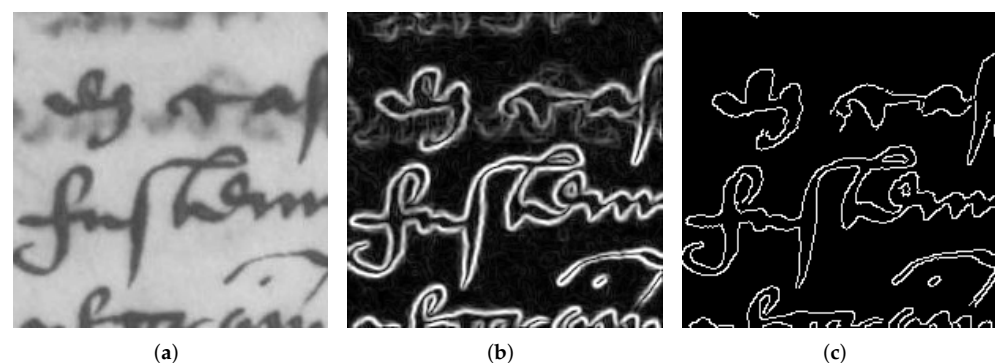


Figure 5. Examples of edge detection on DIBCO2017 dataset. (a) original image; (b) Sobel edge detector; (c) Canny edge detector.

Compared to the simple threshold method, edge detection-based segmentation methods can effectively extract text contour information with fast detection speed. Generally, edge detection methods do not consider the font size, which is an advantage. However, since it relies on calculating pixel gradients within the image itself, it is more sensitive to changes in lighting and can be easily affected by variations in illumination.

Various paradigms are often used in image analysis and processing. Based on the characteristics of the objects, researchers employ statistical methods to classify target pixels. Fuzzy theory defines a fuzzy set and calculates the membership degree of each pixel belonging to each set through fuzzy logic operations. Finally, it classifies pixels based on their membership degree in order to achieve the purpose of segmentation. This process involves obtaining a classifier through a training dataset to distinguish different data, and then using the classifier to predict unknown data. For example, Support Vector Machines (SVMs) can be used to classify image blocks, which is a form of Supervised Learning. Xiong et al. [63] divided a grayscale image into several blocks and enhanced the local contrast of image blocks to roughly extract the foreground. Subsequently, SVMs are employed to classify image blocks into different categories based on statistical information such as region mean, variance, and histogram, thereby determining the optimal global threshold. After thresholding, the image is seamlessly spliced using image seamless splicing technology. Following this, the Canny operator is applied for stroke edge detection, and adaptive local thresholding is used to eliminate noise near stroke edges.

Another approach is to group similar data together, such as K-means, which fall under Unsupervised Learning. Lai et al. [64] proposed two binarization methods based on K-means clustering, one utilizing intensity and the other utilizing color information. After converting the color image to a grayscale image, they applied a median filter to blur the background and used the Sobel operator to emphasize the vertical edges of the text.

Soua et al. [65] proposed a hybrid binarization based on Kmeans method (HBK) and implemented real-time processing of the parallel HBK method in an OCR system. However, K-means is a hard clustering algorithm and cannot process uncertain information. Although K-means is widely used, researchers have proposed more flexible soft clustering algorithms because of its hard clustering nature and sensitivity to noise. Such an algorithm is the fuzzy C-means clustering algorithm, which can accommodate uncertainties related to data points. There are also clustering algorithms such as Possibilistic-Fuzzy C-means (PFCM) [66] and Kernel-based Fuzzy C-Means (KFCM) [67]. Tong et al. [68] combined the Niblack algorithm and the Fuzzy C-Means algorithm (FCM) to propose a camera-based document image binarization algorithm named NFCM. It is expected to address the issue of document image breakage or blurring, preserve the fine details of character strokes, and eliminate glint interference. Furthermore, there are many blur processing algorithms. For instance, Biswas et al. [69] apply a Gaussian filter to the input degraded image files for blur processing. In addition, there is the FuzBin-based binarization method, which Annabestani et al. [70] use to extract text information from document images. They enhance image contrast with fuzzy expert systems (FESs) and then combine FESs with a pixel counting algorithm to obtain a range of threshold values. The middle value is taken as the final threshold. Finally, a method based on mathematical morphology operations, as employed by Gatos et al. [71], is used to enhance blurred stroke information. We summarize these methods in Table 3.

Table 3. Traditional binarization techniques (3).

Classification	Algorithm	Description	Performance
Edge detection	Lu [42]	It detects the text stroke edges based on the local image variation. Combining L1 parametric image gradient and local threshold segmentation.	Suitable for scanned documents with uniform color and texture. Unable to handle situations where the document is skewed or folded.
	T.Lelore [59]	The improved Canny method for edge detection is employed to achieve rough text localization.	This method has relatively low computational costs and can be applied in real-time applications.
	Santhanaprabhu [58]	This method first constructs an adaptive contrast map, and then combines the Sobel edge map to detect the edges of text strokes.	This method has few parameters and can adapt to various degradation types such as uneven illumination and document smear.
	Holambe [60]	Adaptive image contrast combined with Canny's edge map to Edge detection identify stroke-edge pixels.	The approach is low complexity, convenient, and involves fewer parameters.
	Jia [61]	By deforming the Sobel operator to compute the image gradient map. Then, a voting framework is employed to compensate for inaccurate structural symmetry pixels (SSPs).	Due to its adaptive stroke width estimation, this method performs slightly better on printed document images compared to handwritten images.
Fuzzy Logic	Lai [64]	Binarization is performed on images captured by mobile devices using the K-means clustering approach to address degradation issues.	It can reasonably restore Korean text, but the effectiveness of binarized images with fine strokes needs further investigation.
	Tong [68]	Niblack and Fuzzy C-Means (FCM) were used for clustering and to calculate the local thresholds, respectively.	It can effectively preserve strokes and alleviate ghost artifacts. But its performance is unsatisfactory in handling weak strokes.
	Biswas [69]	Blurring of the input degraded file image with Gaussian filter.	Effective for images with a simple background, but may lead to stroke fragmentation when dealing with faint strokes.
	Soua [65]	A method based on hybrid binarization and implemented processing of the parallel binarization based on Kmeans method in an optical character recognition (OCR) system.	Suitable for high-quality printed document images and can be applied in real-time scenarios.
	Annabestani [70]	A global threshold selection method based on fuzzy expert systems (FESs).	It falls within the category of global threshold methods and may require manual adjustment of local threshold for degraded documents. It is more suitable for uniform background.

In addition to the methods mentioned above, there are numerous other techniques available for binarizing document images. Although these methods may not be as widely used as other commonly adopted algorithms, they still hold unique value in practical applications. These include histogram-based methods such as [72–74]; entropy-based methods such as [75]; space binarization-based methods such as [76]; and object property-based methods such as [77], etc. In general, when faced with the task of document image binarization, it is important to comprehensively consider the characteristics and final requirements of the image to choose the most suitable method. Sometimes, combining or layering multiple methods can be an effective approach to improve the binarization of document images. This integrated thinking and practice ensure good results in real situations.

3. Deep Learning Binarization Techniques

Researchers hope to design a model that can receive all image information and address document degradation. A good approach is to design and train a neural network model. Badekas et al. [78] learned from the binarization results generated by various techniques using neural networks. They propose an integrated system for binarizing normal and degraded printed documents to enhance the visualization and recognition of textual characters. While this approach is highly effective for files with complex backgrounds and images, it can also result in lengthy processing times [79]. Deep learning, as a prominent research area in the field of artificial intelligence in recent years, has demonstrated significant potential in the field of document restoration. Therefore, researchers apply deep learning

technology to document image binarization. This method can not only compensate for the limitations of traditional algorithms in handling degraded documents but also offers significant benefits in enhancing the efficiency and accuracy of document processing.

3.1. Based on Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are a deep learning model with a basic structure that includes a convolutional layer, a pooling layer, and a fully connected layer. The convolutional layer is the core component of CNNs. In the field of document image binarization, a convolutional neural network first performs a convolution operation on the input document image to extract the feature information within the image. Then, through the pooling operation, the size of the feature map is reduced to decrease computation, while still preserving important feature information. The collected feature information is then classified in the fully connected layer, resulting in a binary outcome. Usually, researchers combine CNNs with other techniques, such as regularization and data enhancement, to optimize performance and enhance the model's generalization ability. Pastor-Pellicer et al. [80] compared the similarities and differences between Multilayer Perceptrons (MLPs) and introduced the practical application of CNNs in document image binarization. The experiments conducted on the (H)DIBCO [20,22,81,82] and Santgall [83,84] datasets demonstrated that CNNs outperformed MLPs in this task, with particularly notable performance on the Santgall dataset. Various network architectures of CNNs can be utilized. Different from traditional threshold methods, researchers mainly train neural networks to learn degradation and subsequently restore degraded images. In the work by Saddami et al. [85], three deep CNN architectures were compared: Resnet101 [16], Mobilenet V2 [86], and Shufflenet [87] in the task of degradation classification. Shufflenet achieved superior performance in terms of accuracy and computational efficiency. He et al. [88] combined a CNN and the Otsu algorithm to propose an image binarization model called DeepOtsu. Through the automatic feature extraction of deep learning, the threshold selection method in the Otsu algorithm is optimized, resulting in better binarization results. Compared to the Otsu algorithm, the DeepOtsu model can handle more complex image scenes and has a stronger performance against interference factors such as lighting and noise. In a more complex approach, Vo et al. [89] proposed a new supervised binarization method based on a deep supervised network (DSN). The layered DSN architecture is used to learn how to predict text pixels at different feature levels. The network distinguishes text pixels from background noise using higher-level features. The layered architecture helps the proposed approach to retain text strokes more efficiently and provides excellent visual quality. Meng et al. [90] proposed a framework based on deep convolutional neural networks (DCNN). Firstly, the degraded document images are decomposed into spatial pyramid structures by a decomposition network. This network learns character features from images of different scales. A deconvolution network is then used to reconstruct the foreground image from each of these layers in a coarse-to-fine manner.

In 2015, Long et al. [91] proposed Fully Convolutional Networks (FCNs), which for the first time applied deep learning to the field of semantic segmentation. FCN classifies images at the pixel level. It can integrate input image features of any size and then upsample them using deconvolution. The feature is restored to the original input image size, and a label can be generated for each pixel. Tensmeyer [92] described binarization as a task of pixel classification and proposed an algorithm for binarization of low-quality document images and palm leaf manuscript images based on FCN. The FCN algorithm can not only achieve binarization but also recognize and segment different types of objects in the image at the same time. Compared to traditional convolutional neural networks, the advantage of FCN is that it does not limit the size of the image. It also does not require the image to be the same size. The prediction is then realized at the pixel level through the deconvolution output, which produces binary results of the same size as the original image. Secondly, it avoids duplicate storage and computation, making it more efficient. Of course, the shortcomings of FCN are also evident. It is insufficiently sensitive to the details in image

processing, and the results obtained are insufficiently refined. In tasks that require high levels of detail, such as image processing of ancient document images, FCN has potential for further improvement. Ayyalasomayajula et al. [93] proposed an end-to-end structure that combines the FCN and the Primal-Dual network (PD-Net [94]) to address the issue of the foreground category of FCN being either too high or too low in the binarization of document images. The performance and accuracy of the model have been improved. We summarize the document image binarization method based on CNN in Table 4.

Table 4. Deep learning techniques (1).

Algorithm	Description	Performance
Pastor-Pellicer [80]	Describes a practical application of Convolutional Neural Networks (CNNs) in document image binarization tasks.	The method still lags behind advanced approaches, but their work validates the superiority of CNN over Multilayer Perceptrons (MLPs).
Meng [90]	Deep Convolutional Neural Network (DCNN)-based framework. Binarization is employed based on the characteristics or patterns of characters.	It performs well on degraded document images. However, defects may appear in the middle of wide color strokes, and deep stains may not be completely removed.
Vo [89]	A new supervised binarization method based on Deep Supervised Networks (DSNs).	The results retained only a small amount of noise, with few thin or weak strokes missing. Plus, its processing time needs improvement.
He [88]	Combining the Otsu algorithm and CNN results in a method named DeepOtsu.	Enhancing degraded document images before binarization leads to clear and uniform text.
Tensmeyer [92]	Fully Convolutional Network (FCN) based binarization algorithm for low-quality document images.	FCN has certain limitations, and there might be effective features that it cannot learn.
Ayyalasoma-yajula [93]	Combining a FCN and a Primal-Dual Network (PD-Net).	It is suitable for historical document images and transfer learning scenarios.

3.2. Based on Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs) [17] are composed of generator networks and discriminator networks. GANs perform well in binarization, text region detection, and text recognition, among other applications. They have a wide range of applications in image processing, as mentioned in [95–98]. GANs transform the binarization task from a classification problem into an image generation problem. The generator's task is to convert the input grayscale image into a binary image, while the discriminator's task is to judge whether the generated binary image is correct. Through training, the generator can learn how to convert the original document image into a high-quality binarized image, and the discriminator can learn how to accurately distinguish the image generated by the generator from the real binarized image. However, there are still some problems in the training process of GANs, such as unstable training, which requires a significant amount of computing resources and time.

Suh et al. [99] proposed a two-stage GAN for document image binarization with color noise and background removal. In the first stage, the background information is removed, and the color foreground information is extracted to enhance the document image. In the second stage, the binarized image generated by the adversarial network is used to achieve the binarization of the document image. Bhunia et al. [100] built a Texture Augmentation Network (TANet) by introducing adversarial learning to transfer the texture elements of degraded reference document images into a clean binarized image. This method has various noise texture versions of the same text content and expands the training set. Bhowmik et al. [101] drew inspiration from game theory and utilized unsupervised learning in their approach. Specifically, they use the K-means clustering method to classify pixels into foreground and background. The quality of binarization is enhanced through pre-processing and post-processing. Kumar et al. [102] optimized Bhunia's algorithm by introducing a joint discriminator to combine TANet and the unsupervised document binarization network (UDBNet). This enhancement addresses dataset bias, aiming to achieve improved performance on actual degraded images. To generate more challenging

adversarial samples for UDBNet training, researchers utilized an Adversarial Texture Augmentation Network (ATANet) to create a pseudo image pair. Konwer et al. [103] used GAN to remove staff line to achieve binarization in the pre-processing step of optical music recognition. Zhao et al. [104] introduced conditional generation adversarial networks (cGANs) [105] to solve the problem of multi-scale information composition in binary tasks. Souibgui et al. [106] used cGANs to propose a pix2pix framework called Document Enhancement Generative Adversarial Network (DE-GAN) to restore severely degraded document images. The discriminator inputs the degraded image and the Ground Truth (GT), and it compels the generator to generate an output that is indistinguishable from the GT. After the training is completed, the discriminator becomes unnecessary, and only the generator network is used to enhance the degraded image. Figure 6 shows the binarization result of this method, which still retains some background noise and slightly widens the text strokes.

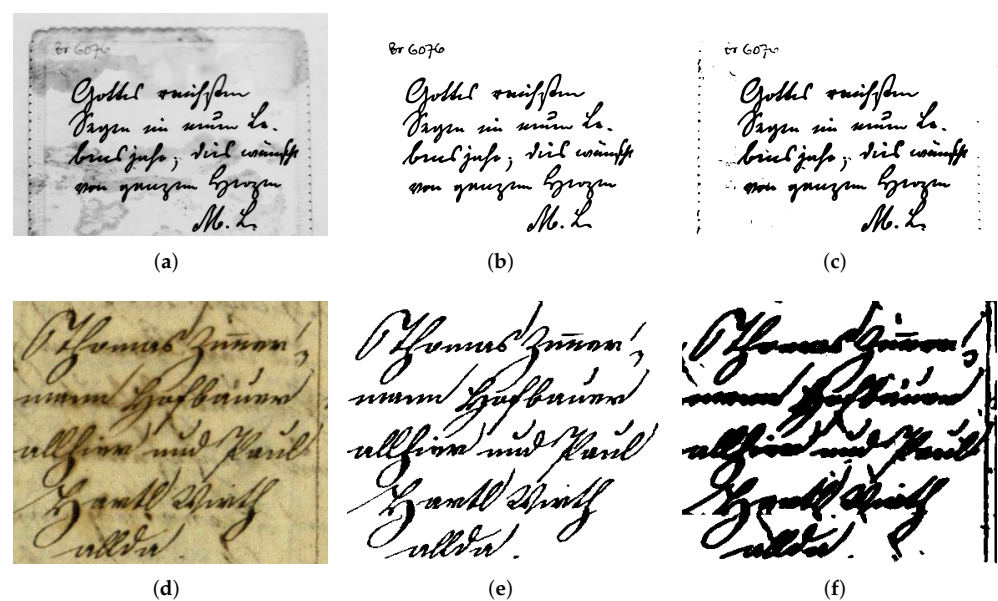


Figure 6. Examples of DE-GAN's [106] results on HDIBCO2016 dataset. (a) original image; (b) ground truth image of (a); (c) binarization result of (a); (d) original image; (e) ground truth image of (d); (f) binarization result of (d).

R. De et al. [107] propose a Dual Discriminator Generative Adversarial Network (DD-GAN) that utilizes Focal Loss as the generator loss. The model uses a network of two discriminators to capture information. The global discriminator is responsible for higher-level image features, such as image background and texture, while the local discriminator focuses on lower-level features like text strokes. Additionally, the model employs focal loss which is used to solve the issue of class imbalance among pixels. Rajesh et al. [108] argue that while most existing technologies concentrate on pixel images as input, they may not yield satisfactory outcomes when processing compressed images that require complete decompression. Therefore, Rajesh applied DD-GAN and proposed the direct use of JPEG for compressing document images to achieve binarization. Lin et al. [109] proposed a three-stage approach to enhance and binarize degraded color document images by using discrete wavelet transform (DWT) and GANs. This general model approach can be trained with different wavelet transforms and neural networks. This method can be effectively applied to the degraded color document image binarization task. We summarize the document image binarization method based on GANs in Table 5.

Table 5. Deep learning techniques (2).

Algorithm	Description	Performance
Bhunja [100]	A texture enhancement network is constructed by introducing an adversarial learning approach.	It can provide better visual quality but might compromise stroke edge details, resulting in noise within thick strokes.
Zhao [104]	The conditional GANs (cGANs) are introduced to solve the multi-scale information combination problem in binarization tasks.	It performs well under various degradation scenarios, but there are minor noise dots in the middle of the strokes.
Bhowmik [101]	They proposed a document image binarization method based on game theory, using the K-means clustering algorithm for pixel classification.	It effectively eliminates artifacts while preserving some boundary pixels of stamps and stains.
Kumar [102]	The joint discriminator combines the Texture Augmentation Network (TANet) and unsupervised document binarization network (UDBNet) to address the issue of dataset bias.	In handwritten document images, stroke omissions and background noise may occur.
Suh [99]	A two-stage color document image enhancement and binarization method using GANs. The study focuses on the issue of multi-color degradation.	It can solve the problem of document degradation, including postmarks, bleed-through, uneven contrast, and so on.
Souibgui [106]	A conditional generative adversarial network for document enhancement named DE-GAN.	It removes most background imprints, but the method also preserves some noise, as well as traces like paper creases.
R.De [107]	A Dual Discriminator GAN (DD-GAN) is proposed. It utilizes both global and local information about pixel distribution.	The overall performance is good, but the performance in HDIBCO2018 is poor, mainly due to the inability to distinguish background pixels that do not belong to the manuscript.
Rajesh [108]	The proposed model directly binarizes document images in their compressed form without the need for decompression.	The most important point of this method is its ability to significantly enhance computing efficiency and save storage space.
Fathallah [110]	A historical document image enhancement model is proposed based on GANs.	Some stains, such as watermarks, can be removed, but the resulting text strokes will be thinner, which may lead to the loss of some details.

3.3. Based on Attention Mechanism

The Attention Mechanism is a special structure in machine learning that simulates the selective perception of certain information by human attention. It automatically selects the most important part of the input data, reducing the impact of noise. Additionally, it can be used to enhance the expression and generalization ability of the network. For example, in document image processing, the attention mechanism can learn to calculate the weight coefficients of different areas. This allows for more attention to be paid to the text or background areas. Guo et al. [111] proposed a novel Multi-scale Multi-attention Network (MsMa-Net) for the fresh moiré document image binarization task. Peng et al. [112] proposed a deep learning framework for inferring the probability of a text region using a multi-resolution attention model. This probability is then fed into a convolutional conditional random field (ConvCRF) to obtain a final binarized document image. The author uses a neural network to learn the features of degraded document images and employs ConvCRF to infer the relationship between text areas and the background. The author claims that this approach can result in stronger generalization ability.

The encoder-decoder structure is a common model structure in deep learning. The encoder converts the input data into an intermediate value that captures the key characteristics of the input data. The decoder receives the median value from the encoder and uses it to generate an output, such as a pixel value for an image or a sequence of words for text. In Natural Language Processing (NLP), common encoder-decoder structures include Seq2Seq and Transformer. In the field of image processing, the commonly used encoder-decoder structures are U-Net and VGG. In encoder-decoder structures, attention mechanisms are often used, as seen in the implementation of U-Net. This process helps extract useful features from the original document image and generate an accurately binarized document image.

In a document image binarization task, the encoder typically converts the document image into a sequence of vectors to capture the key features of the document image. The decoder then uses this sequence to generate a binarized document image. The U-Net was proposed in 2015 and was initially applied to image segmentation tasks in the biomedical field. It uses an encoder-decoder structure, in which the encoder is responsible for extracting features, while the decoder restores the image to its original resolution. Different from FCN's feature addition mechanism, U-Net concatenates the up-sampled and down-sampled feature maps by skip connections to preserve more dimension and location information. This improves the segmentation effect of the network. Therefore, the structure of U-Net is highly suitable for document image segmentation. Bezmaternykh et al. [113] used the U-Net architecture to propose a CNN-based method called U-Net-bin, which won first place in the DIBCO'17 competition. Furthermore, they argue that the success of binarization is not crucial in the form of Chinese or English characters. Xiao et al. [55] also used U-Net architecture as the foundation to propose a method for document binarization that combines local and global features. Based on the attentional U-Net, Zhao et al. [114] proposed a binarization method for historical Tibetan document images. In this method, the input image is unsampled twice during the inference stage to alleviate pseudo-touching. Ke Ma et al. [115] combined U-Net and Transformer models to perform end-to-end training for geometric correction and binarization of document images. They used a stacked U-Net with intermediate supervision for this purpose.

Peng et al. [116] proposed a convolutional encoder-decoder model specifically designed for the binarization of document images. The encoder is constructed by stacking convolutional layers to learn the features of the middle layer of the document image. The low-resolution representation is then mapped to the original size using a decoder to generate the final binarized image. Souibgui et al. [117] adopted the Vision Transformers model for the binarization of document images and named it DocEnTr. The model captures high-level global remote dependencies through a self-attention mechanism and outputs binarized images of documents in an end-to-end manner.

Chaurasia et al. [118] proposed a network architecture called LinkNet, which drew inspiration from the U-Net model and adopted an Encoder-Decoder structure to create a lightweight network capable of real-time segmentation. Xiong et al. [119] proposed an improved semantic segmentation model called DP-LinkNet, which is based on the LinkNet and D-LinkNet [120] models. They introduced a Hybrid Dilated Convolution (HDC) module in the middle of the architecture to increase the receptive field and enhance the network's ability to capture details and textures in images. The Spatial Pyramid Pooling (SPP) has also been introduced to improve the perception of features at different scales. The experimental results show that the proposed method performs well on document images with noise, such as stains and imprints, and achieves excellent speed and accuracy. As shown in Figure 7, the two models exhibit little visual difference. Both effectively extract text, with only minor shortcomings in the details. Specifically, there are traces of discontinuity in fine strokes, and the edges of blurry strokes are not as sharp. Despite this, they have achieved a satisfactory outcome.

In general, U-Net can assist in document image processing by preserving text information and eliminating background noise. When combined with an attention mechanism, it can effectively improve the efficiency of document image processing. In the document image binarization, the traditional threshold segmentation method and morphological operation method can be combined to optimize the binarization result. Additionally, they can also be used to post-process the output of U-Net for further optimization. For example, morphological operations can be used to perform dilation and erosion operations, which help in removing noise and small fragmented areas. Additionally, threshold segmentation can also be used to extract more detailed information. At the same time, U-Net can also be combined with other deep learning models using multi-task learning. This approach allows for simultaneous text detection and binarization, thereby improving the efficiency of the entire document processing. In many tasks, these neural network algorithms can

achieve high accuracy in binarization. Most of them do not require the pre-processing of the document image. However, due to the complexity of the neural network, the calculations may take some time. We summarize the document image binarization method based on the attention mechanism in Table 6.

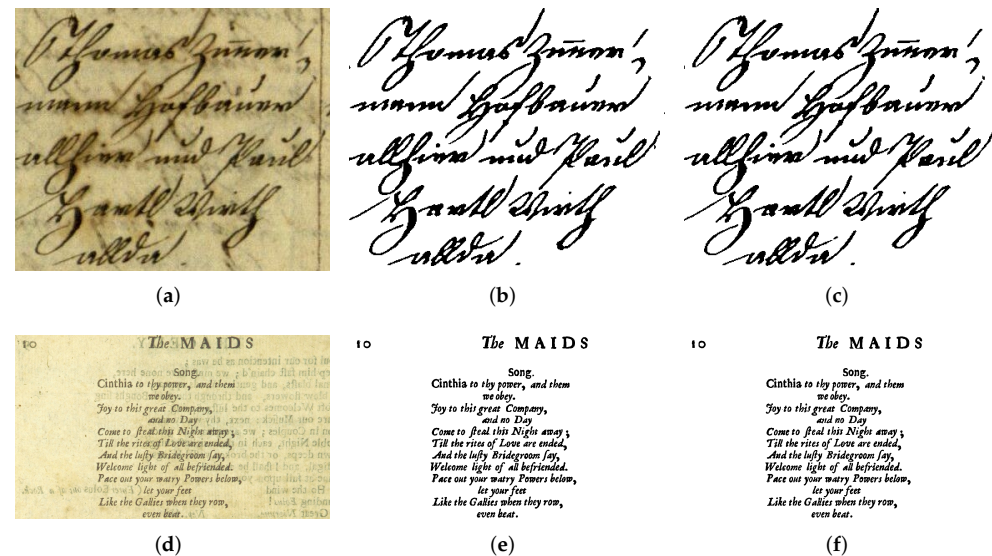


Figure 7. Binarization results of handwritten document image from HDIBCO2016 dataset and printed document image from DIBCO2017 dataset. (a) original image; (b) LinkNet's result of (a); (c) DP-LinkNet's result of (a); (d) original image; (e) LinkNet's result of (d); (f) DP-LinkNet's result of (d).

Table 6. Deep learning techniques (3).

Algorithm	Description	Performance
Peng [116]	The probabilities of text regions are inferred from a multi-resolution attention model, which is then fed into a ConvCRF.	It performs well in document images with contrast variations and exhibits good generalization capability.
Ke Ma [115]	A hybrid model based on U-Net and Transformer for flattening and correcting distorted deformations in document images.	It significantly restores distorted document images, which is highly beneficial for the binarization of degraded documents.
Peng [112]	It extrapolates the probability of the text region using a multi-resolution attention model and feeds it into ConvCRF to obtain the final binarized document image.	It can effectively remove degraded stain marks, but it is easy to identify the page boundary.
Bezmaternykh [113]	It uses U-Net architecture to achieve more accurate historical document images binarization.	It can exhibit excellent results for document images, and this method secured the first position in the DIBCO '17 competition.
Xiao [55]	It utilizes a combination of global and local branches, adopting U-Net as its architecture, and finally produces the ultimate binarization result through logical operator fusion.	Comparative experiments indicate that the combined approach can enhance the discriminative performance between text and background.
Guo [111]	A Multiscale Multi-Attention Network (MsMa-Net).	It can eliminate the majority of background noise, but it has some limitations as the edges of strokes may not be smooth enough.
Zhao [114]	A U-Net-based binarization method provides a solution to the degradation issues in historical Tibetan document images, particularly pseudo-touching strokes.	This method achieves the best results in enlarged Tibetan images, but it may produce holes in the strokes of faint characters.
Xiong [119]	Based on the LinkNet architecture, the model integrates the Hybrid Dilated Convolution (HDC) and Spatial Pyramid Pooling (SPP) modules between the encoder and the decoder.	It can effectively remove background stains, restore text information, and is much faster than most CNN-based methods.

4. Results

Different techniques are utilized for evaluating document image binarization. Firstly, the most common approach involves human visual observation [121], but this intuitive method lacks quantitative analysis. Secondly, an end-to-end approach may be employed, such as using OCR performance as a reference [45], but the results obtained through this method may also reflect the influence of other image tasks. Unsupervised metric methods typically assess the quality of binarization by analyzing properties of image segmentation, such as those based on gray-intensity variances [122]. However, the method may lead to misleading results. Therefore, for the evaluation criteria of binarization in both handwritten and printed document images, pixel-based binarization evaluation methods are widely employed. This paper utilizes several well-known evaluation metrics commonly used in DIBCO competitions for quality assessment.

4.1. Performance Measures

4.1.1. PSNR

PSNR, which stands for Peak Signal-to-Noise Ratio, is used to measure the similarity between the original image and the processed image. This index is calculated by determining the Mean Square Error (MSE) between the pixel values of the two images and converting it into decibel (dB) units. To quantify the relative error between two images, the larger the value, the higher the similarity, the smaller the relative error, the better the image quality. In document image binarization, the resulting image after binarization consists of only black and white pixel values. This binarization result can be considered as a compressed or distorted image. The PSNR value can be calculated by calculating the MSE between the binarized image and the original image.

$$PSNR = 10 \log \left(\frac{MAXI^2}{MSE} \right), \quad (5)$$

$$MSE = \frac{\sum_{x=1}^M \sum_{y=1}^N (I_{bin}(x, y) - I_{GT}(x, y))^2}{MN}. \quad (6)$$

where $MAXI$ denotes the maximum image pixel value, usually 255, while MSE denotes the mean square error between the compressed and original images. $I_{bin}(x, y)$ denotes the image pixel value after binarization and $I_{GT}(x, y)$ denotes the image pixel value of the reference image (Ground Truth).

4.1.2. F-Measure

F-measure (FM) is a metric that combines precision (P) and recall (R), and it is the harmonic mean of the two values. The following equation can express it as follows:

$$FM = \frac{(1 + \beta^2)PR}{\beta^2P + R}, \quad (7)$$

where β is a weighting factor, generally taking a value of 1, indicating that Precision and Recall are equally important.

When $\beta = 1$, which is also called the F1-Score. Precision indicates the proportion of pixels that are binarized as foreground and truly belong to the foreground, while Recall indicates the proportion of pixels that truly belong to the foreground and are correctly binarized as foreground. They are defined by the following parameters: True Positive (TP) indicates the number of pixels correctly classified as foreground; False Positive (FP) indicates the number of pixels incorrectly classified as foreground; and False Negative (FN) indicates the number of pixels incorrectly classified as background.

$$FM = \frac{2PR}{P + R}, P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN}. \quad (8)$$

In document images, there are typically more background pixels than foreground pixels. The F-measure can penalize methods that produce disproportionate false positives or false negatives.

4.1.3. Pseudo F-Measure

Pseudo F-measure (F_{ps}) [123], an improved algorithm for F-measure, is mainly used for the evaluation of binarization, which is calculated as follows:

$$F_{ps} = \frac{2 \cdot R_{ps} \cdot P_{ps}}{R_{ps} + P_{ps}}. \quad (9)$$

F_{ps} introduces pseudo-recall (R_{ps}) and pseudo-precision (P_{ps}), considering local stroke width and the distance from the contour of the ground truth text, thereby more effectively capturing binarization performance.

4.1.4. DRD

Distance Reciprocal Distortion (DRD), is a metric for image quality evaluation (cf. [124]), mainly used to measure the sharpness and contrast of the image. It has a good correlation with the error perception of human visual detection, so DRD is used to measure the visual distortion in document images binarization.

$$DRD = \frac{\sum_{k=1}^N DRD_k}{NUBN}, \quad (10)$$

where DRD_k denotes the distortion of the k th flipped pixel, and $NUBN$ denotes the number of non-uniform color blocks in the reference image (GT).

4.2. Experimental Result

In this subsection, we compare the quality results of 25 techniques using the dataset H-DIBCO2016 [23] from the Handwriting Document Image Binarization Competition. The first part involves traditional techniques, utilizing threshold algorithms from commonly used Python libraries such as OpenCV and scikit-image for quality assessment. The evaluation results for the second part, which are based on deep learning techniques, are extracted from the data provided in the respective original papers. The omission of results for other techniques is due to the unavailability of implementation source code and limited details provided in the original papers. These constraints influenced the selection of methods included in the quality assessment results.

As observed in Table 7, it can be inferred that the performance of methods based on deep learning models tends to surpass that of traditional threshold-based binarization methods, especially in the context of the HDIBCO2016 dataset. According to the results, Kumar et al. [102] proposed that the unsupervised document binarization network had the best performance in terms of three indicators: F-Measure, PSNR, and DRD. In the Pseudo F-measure index, the best performance was achieved by the three-stage binarization of color document images proposed by Lin et al. [109]. Among threshold-based methods, the Otsu algorithm [31], Wolf algorithm [49], and Gatots algorithm [71] specifically demonstrate better comprehensive evaluation indicators. On the other hand, algorithms based on deep learning models exhibit better overall performance. Notably, the model algorithms proposed by He [88], Zhao [104], and Peng [112] have shown promising results.

Table 7. Quality evaluation results of different algorithms on the HDIBCO2016 dataset.

Classification	Algorithm	F-Measure	$F_p s$	PSNR	DRD
Traditional based	Otsu [31]	86.59	89.92	17.79	6.89
	Niblack [32]	63.10	63.64	12.47	31.48
	Bernsen [48]	71.73	75.87	13.66	25.41
	Sauvola [34]	81.27	83.91	15.36	17.92
	Wolf [49]	86.36	90.46	17.11	9.01
	Gatos [45]	86.59	89.09	17.47	7.01
	NICK [36]	80.86	82.65	15.15	19.24
	Su [46]	74.41	87.03	15.52	12.55
	Hadjadj [41]	83.68	87.03	16.20	13.05
	T.R.Singh [38]	84.47	87.56	16.99	9.73
	Bataineh [37]	83.27	84.43	16.02	15.61
Deep learning based	WAN [26]	75.14	76.06	13.61	30.33
	Vo [89]	90.10	93.57	19.01	3.58
	Tensmeyer [92]	89.52	93.76	18.67	3.76
	He [88]	91.4	94.3	19.60	2.9
	Meng [90]	89.90 ± 4.55	-	18.79 ± 3.36	-
	Ayyalasomayajula [93]	90.18	-	18.99	3.61
	Zhao [104]	91.66	94.58	19.64	2.82
	Bhowmik [101]	91.15	-	19.18	3.20
	Kumar [102]	93.4	96.2	20.1	2.2
	Lin [109]	91.46	96.32	19.66	2.94
	R.De [107]	89.98	95.23	18.83	3.61
	Xiao [55]	90.77	94.21	19.33	3.11
	Peng [112]	91.68	-	19.59	2.93
	Peng [116]	88.07 ± 4.86	-	18.13 ± 3.13	-

Bold font indicates the best result.

5. Conclusions

Document image binarization is a complex and multi-level process. Due to the various types and degrees of damage to document images, the processing emphasis varies. Because of this, there is no universal binarization method that works for all types of document images.

This paper provides a detailed overview of approximately sixty document image binarization methods, evaluating the quality of twenty-five of them using the H-DIBCO2018 dataset. Based on the evaluation, traditional binarization algorithms perform well in handling document images with simple backgrounds but show limited effectiveness when dealing with images with complex backgrounds, especially those containing mixed noise or severe contamination. Deep learning methods, by employing pixel-level image segmentation, produce promising results while also addressing complex details within the images. However, it is worth noting that these methods may require more computational resources and time.

When choosing specific methods, one must consider the task requirements. This may include degradation issues in the document, the type of document (handwritten or printed), and the availability of sufficient computational resources. Additionally, the type of degradation in the document should be taken into account when choosing the most suitable method.

Given the potential variations in characters and numbers in document images from different regions, future research should focus on enhancing the cross-language capabilities of algorithms to effectively handle stroke characteristics of different texts. Depending on the requirements of different scenarios, future studies can design algorithms that combine knowledge from various fields, such as image enhancement and restoration, or fine-tuning existing advanced algorithms.

Due to the complex data characteristics of degraded document images, each image may exhibit different types and levels of damage or degradation. Therefore, the primary challenge in training current networks is the lack of ground truth, which leads to insufficient datasets. To address this challenge, future research can delve into the application of unsupervised techniques. We also plan to continue collecting and exploring the application of unsupervised techniques in document image binarization.

Additionally, to comprehensively assess the performance of document image binarization methods, it is necessary to further explore and introduce new evaluation metrics. While existing metrics are widely used, they may inherently possess inaccuracies, prompting the search for more precise and comprehensive performance measures. We are also committed to evaluating binarization techniques by introducing additional high-quality metrics and utilizing diverse datasets. In future research, we will delve into exploring and evaluating a select few highly practical techniques.

Author Contributions: Conceptualization, Z.Y. and S.Z.; methodology, Z.Y.; software, Z.Y.; validation, Y.Z., J.H. and J.S.; formal analysis, Z.Y.; investigation, Z.Y. and Y.Z.; resources, S.Z.; data curation, Z.Y. and Y.Z.; writing—original draft preparation, Z.Y.; writing—review and editing, S.Z.; visualization, J.H. and J.S.; supervision, S.Z.; project administration, S.Z.; funding acquisition, S.Z. All authors have read and agreed to the published version of the manuscript.

Funding: Natural Science Foundation of Fujian Province of China (General Program): 2022J011273; Educational Teaching Reform Research Project of Xiamen University of Technology in 2022: JG202209.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The deep learning-based data in the results section comes from the experimental data of the original author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Gatos, B.; Pratikakis, I.; Kepene, K.; Perantonis, S.J. Text Detection in Indoor/Outdoor Scene Images. 2005. Available online: https://www.researchgate.net/publication/253135219_Text_Detection_in_IndoorOutdoor_Scene_Images (accessed on 26 March 2024).
2. Pan, Y.F.; Hou, X.; Liu, C.L. Text Localization in Natural Scene Images Based on Conditional Random Field. In Proceedings of the 2009 10th International Conference on Document Analysis and Recognition, Barcelona, Spain, 26–29 July 2009; pp. 6–10.
3. Liao, M.; Wan, Z.; Yao, C.; Chen, K.; Bai, X. Real-time Scene Text Detection with Differentiable Binarization. *arXiv* **2019**, arXiv:1911.08947.
4. Kamnitsas, K.; Ledig, C.; Newcombe, V.F.J.; Simpson, J.P.; Kane, A.D.; Menon, D.K.; Rueckert, D.; Glocker, B. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* **2016**, *36*, 61–78. [CrossRef] [PubMed]
5. Atia, N.; Benzaoui, A.; Jacques, S.; Hamiane, M.; Kourd, K.E.; Bouakaz, A.; Ouahabi, A. Particle Swarm Optimization and Two-Way Fixed-Effects Analysis of Variance for Efficient Brain Tumor Segmentation. *Cancers* **2022**, *14*, 4399. [CrossRef] [PubMed]
6. Gupta, M.R.; Jacobson, N.P.; Garcia, E.K. OCR binarization and image pre-processing for searching historical documents. *Pattern Recognit.* **2007**, *40*, 389–397. [CrossRef]
7. Murdock, M.; Reid, S.; Hamilton, B.; Reese, J.W. ICDAR 2015 competition on text line detection in historical documents. In Proceedings of the 2015 13th International Conference on Document Analysis and Recognition (ICDAR), Tunis, Tunisia, 23–26 August 2015; pp. 1171–1175.
8. Kumar, G.; Bhatia, P.K. A Detailed Review of Feature Extraction in Image Processing Systems. In Proceedings of the 2014 Fourth International Conference on Advanced Computing & Communication Technologies, Rohtak, India, 8–9 February 2014; pp. 5–12.
9. Marques, O. Morphological Image Processing. 2011. Available online: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118093467.ch13> (accessed on 26 March 2024).
10. Weszka, J.S.; Rosenfeld, A. Threshold Evaluation Techniques. *IEEE Trans. Syst. Man Cybern.* **1978**, *8*, 622–629. [CrossRef]
11. Weszka, J.S. A survey of threshold selection techniques. *Comput. Graph. Image Process.* **1978**, *7*, 259–265. [CrossRef]
12. Sahoo, P.K.; Soltani, S.; Wong, A.K.C. A survey of thresholding techniques. *Comput. Vis. Graph. Image Process.* **1988**, *41*, 233–260. [CrossRef]
13. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
14. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2012**, *60*, 84–90. [CrossRef]
15. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
16. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2015; pp. 770–778.
17. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *arXiv* **2014**, arXiv:1406.2661.

18. Radford, A.; Metz, L.; Chintala, S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv* **2015**, arXiv:1511.06434.
19. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Proceedings, Part III 18; Springer: Berlin/Heidelberg, Germany, 2015.
20. Gatos, B.; Ntirogiannis, K.; Pratikakis, I. DIBCO 2009: document image binarization contest. *Int. J. Doc. Anal. Recognit. (IJ DAR)* **2011**, *14*, 35–44. [\[CrossRef\]](#)
21. Pratikakis, I.; Gatos, B.; Ntirogiannis, K. ICDAR 2011 Document Image Binarization Contest (DIBCO 2011). In Proceedings of the 2011 International Conference on Document Analysis and Recognition, Beijing, China, 18–21 September 2011; pp. 1506–1510.
22. Pratikakis, I.; Gatos, B.; Ntirogiannis, K. ICDAR 2013 Document Image Binarization Contest (DIBCO 2013). In Proceedings of the 2013 12th International Conference on Document Analysis and Recognition, Washington, DC, USA, 25–28 August 2013; pp. 1471–1476.
23. Pratikakis, I.; Zagoris, K.; Barlas, G.; Gatos, B. ICFHR2016 Handwritten Document Image Binarization Contest (H-DIBCO 2016). In Proceedings of the 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), Shenzhen, China, 23–26 October 2016; pp. 619–623.
24. Pratikakis, I.; Zagoris, K.; Barlas, G.; Gatos, B. ICDAR2017 Competition on Document Image Binarization (DIBCO 2017). In Proceedings of the 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, Japa, 9–15 Novemver 2017; pp. 1395–1403.
25. Bhowmik, S. Document Image Binarization. In *Document Layout Analysis*; Springer Nature Singapore: Singapore, 2023. Available online: https://link.springer.com/chapter/10.1007/978-981-99-4277-0_2 (accessed on 26 March 2024).
26. Mustafa, W.A.; Kader, M.M.M.A. Binarization of Document Image Using Optimum Threshold Modification. *J. Phys. Conf. Ser.* **2018**, *1019*, 012022. [\[CrossRef\]](#)
27. Patil, P. Survey on document image binarization. *Int. J. Adv. Res. Ideas Innov. Technol.* **2019**, *5*, 273–275.
28. Sezgin, M.; Sankur, B. Survey over image thresholding techniques and quantitative performance evaluation. *J. Electron. Imaging* **2004**, *13*, 146–168.
29. Ismail, S.M.; Abdullah, S.N.H.S.; Fauzi, F. Statistical Binarization Techniques for Document Image Analysis. *J. Comput. Sci.* **2018**, *14*, 23–36. [\[CrossRef\]](#)
30. Saxena, L.P. Niblack’s binarization method and its modifications to real-time applications: A review. *Artif. Intell. Rev.* **2019**, *51*, 673–705. [\[CrossRef\]](#)
31. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [\[CrossRef\]](#)
32. Niblack, W. *An Introduction to Digital Image Processing*; Prentice-Hall International: Englewood Cliffs, NJ, USA, 1986. Available online: <https://archive.org/details/introductiontodi0000nibl> (accessed on 26 March 2024).
33. Trier, Ø.D.; Jain, A.K. Goal-Directed Evaluation of Binarization Methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **1995**, *17*, 1191–1201. [\[CrossRef\]](#)
34. Sauvola, J.J.; Seppänen, T.; Haapakoski, S.; Pietikäinen, M. Adaptive document binarization. In Proceedings of the Fourth International Conference on Document Analysis and Recognition, Ulm, Germany, 18–20 August 1997; Voume 1, pp. 147–152.
35. Lazzara, G.; Géraud, T. Efficient Multiscale Sauvola’s Binarization. *Int. J. Doc. Anal. Recognit. (IJ DAR)* **2014**, *17*, 105–123. [\[CrossRef\]](#)
36. Khurshid, K.; Siddiqi, I.; Faure, C.; Vincent, N. Comparison of Niblack inspired binarization methods for ancient documents. In *Document Recognition and Retrieval XVI*; SPIE: Bellingham, WA, USA, 2009.
37. Bataineh, B.; Abdullah, S.N.H.S.; Omar, K.B. An adaptive local binarization method for document images based on a novel thresholding method and dynamic windows. *Pattern Recognit. Lett.* **2011**, *32*, 1805–1813. [\[CrossRef\]](#)
38. Singh, T.R.; Roy, S.; Singh, O.I.; Sinam, T.; Singh, K.M. A New Local Adaptive Thresholding Technique in Binarization. *arXiv* **2012**, arXiv:1201.5227.
39. Chaki, N.; Shaikh, S.H.; Saeed, K. A Comprehensive Survey on Image Binarization Techniques. In *Exploring Image Binarization Techniques*; Springer: New Delhi, India, 2014. [\[CrossRef\]](#)
40. He, J.; Do, Q.; Downton, A.C.; Kim, J.H. A comparison of binarization methods for historical archive documents. In Proceedings of the Eighth International Conference on Document Analysis and Recognition (ICDAR’05), Seoul, Republic of Korea, 31 August 31–1 September 2005; Volume 1, pp. 538–542.
41. Hadjadj, Z.; Meziane, A.; Cherfa, Y.; Cheriet, M.; Setitra, I. ISauvola: Improved Sauvola’s Algorithm for Document Image Binarization. In Proceedings of the International Conference on Image Analysis and Recognition, Niagara Falls, ON, Canada, 22–24 July 2015.
42. Lu, S.; Su, B.; Tan, C.L. Document image binarization using background estimation and stroke edges. *Int. J. Doc. Anal. Recognit. (IJ DAR)* **2010**, *13*, 303–314. [\[CrossRef\]](#)
43. Mieloch, K.; Mihăilescu, P.; Munk, A. Dynamic threshold using polynomial surface regression with application to the binarization of fingerprints. In *SPIE Defense + Commercial Sensing*; SPIE: Bellingham, WA, USA, 2005.
44. Moghaddam, R.F.; Cheriet, M. RSLDI: Restoration of single-sided low-quality document images. *Pattern Recognit.* **2009**, *42*, 3355–3364. [\[CrossRef\]](#)

45. Gatos, B.; Pratikakis, I.; Perantonis, S.J. Adaptive degraded document image binarization. *Pattern Recognit.* **2006**, *39*, 317–327. [[CrossRef](#)]
46. Su, B.; Lu, S.; Tan, C.L. Binarization of historical document images using the local maximum and minimum. In Proceedings of the International Workshop on Document Analysis Systems, Boston, MA, USA, 9–11 June 2010.
47. Su, B.; Lu, S.; Tan, C.L. Robust Document Image Binarization Technique for Degraded Document Images. *IEEE Trans. Image Process.* **2013**, *22*, 1408–1417.
48. Bernsen, J. Dynamic thresholding of grey-level images. In Proceedings of the Eighth International Conference on Pattern Recognition, Paris, France, 27–31 October 1986.
49. Wolf, C.; Jolion, J.M. Extraction and recognition of artificial text in multimedia documents. *Form. Pattern Anal. Appl.* **2003**, *6*, 309–326. [[CrossRef](#)]
50. Yang, Y. OCR Oriented Binarization Method of Document Image. In Proceedings of the 2008 Congress on Image and Signal Processing, Sanya, China, 27–30 May 2008; Volume 4, pp. 622–625.
51. Zemouri, E.T.; Chibani, Y.; Brik, Y. Enhancement of Historical Document Images by Combining Global and Local Binarization Technique. *Int. J. Inf. Eng. Electron. Bus.* **2014**, *4*, 1. [[CrossRef](#)]
52. Chaudhary, P.; Ambedkar, B. An effective and robust technique for the binarization of degraded document images. *Int. J. Res. Eng. Technol.* **2014**, *03*, 140–145.
53. Ntirogiannis, K.; Gatos, B.; Pratikakis, I. A combined approach for the binarization of handwritten document images. *Pattern Recognit. Lett.* **2014**, *35*, 3–15. [[CrossRef](#)]
54. Liang, Y.; Lin, Z.; Sun, L.; Cao, J. Document image binarization via optimized hybrid thresholding. In Proceedings of the 2017 IEEE International Symposium on Circuits and Systems (ISCAS), Baltimore, MD, USA, 28–31 May 2017; pp. 1–4.
55. Xiao, H.; Lin, L.; Rong, L.; Chengshen, X.; Ye, M. Binarization of degraded document images with global-local U-Nets. *Optik* **2020**, *203*, 164025.
56. Saddami, K.; Arnia, F.; Away, Y.; Munadi, K. Kombinasi Metode Nilai Ambang Lokal dan Global untuk Restorasi Dokumen Jawi Kuno. *J. Teknol. Inf. Dan Ilmu Komput.* **2020**, *7*, 163–170 [[CrossRef](#)]
57. Ranjitha, P.; Shreelakshmi, T.D. A Hybrid Ostu based Niblack Binarization for Degraded Image Documents. In Proceedings of the 2021 2nd International Conference for Emerging Technology (INCET), Belagavi, India, 21–23 May 2021; pp. 1–7.
58. Santhanaprabhu, G.; Karthick, B.; Srinivasan, P.; Vignesh, R.K.; Sureka, K. Extraction and Document Image Binarization Using Sobel Edge Detection. *J. Eng. Res. Appl.* **2014**, *4*, 15–21.
59. Lelore, T.; Bouchara, F. FAIR: A Fast Algorithm for Document Image Restoration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2039–2048. [[CrossRef](#)] [[PubMed](#)]
60. Holambe, S.N.; Shinde, U.B.; Choudhari, B.S. Image Binarization for Degraded Document Images. *Int. J. Comput. Appl.* **2015**, *128*, 38–43.
61. Jia, F.; Shi, C.; He, K.; Wang, C.; Xiao, B. Document Image Binarization Using Structural Symmetry of Strokes. In Proceedings of the 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), Shenzhen, China, 23–26 October 2016; pp. 411–416.
62. Hadjadj, Z.; Cheriet, M.; Meziane, A.; Cherfa, Y. A new efficient binarization method: Application to degraded historical document images. *Signal Image Video Process.* **2017**, *11*, 1155–1162. [[CrossRef](#)]
63. Xiong, W.; Xu, J.; Zijie, X.; Juan, W.L.; Min, L. Degraded historical document image binarization using local features and support vector machine (SVM). *Optik* **2018**, *164*, 218–223. [[CrossRef](#)]
64. Lai, A.N.; Lee, G. Binarization by Local K-means Clustering for Korean Text Extraction. In Proceedings of the 2008 IEEE International Symposium on Signal Processing and Information Technology, Vancouver, BC, Canada, 27–30 August 2006; IEEE: New York, NY, USA, 2008; pp. 117–122.
65. Soua, M.; Kachouri, R.; Akil, M. GPU parallel implementation of the new hybrid binarization based on Kmeans method (HBK). *J. Real-Time Image Process.* **2018**, *14*, 363–377. [[CrossRef](#)]
66. Pal, N.R.; Pal, K.; Keller, J.M.; Bezdek, J.C. A possibilistic fuzzy c-means clustering algorithm. *IEEE Trans. Fuzzy Syst.* **2005**, *13*, 517–530. [[CrossRef](#)]
67. Farahmand, A.; Sarrafzadeh, H.; Shanbehzadeh, J. Noise removal and binarization of scanned document images using clustering of features. In Proceedings of the International MultiConference of Engineers and Computer Scientists 2017 Vol I, IMECS 2017, Hong Kong, China, 15–17 March 2017.
68. Tong, L.; Chen, K.; Zhang, Y.; Fu, X.L.; Duan, J. Document Image Binarization Based on NFCM. In Proceedings of the 2009 2nd International Congress on Image and Signal Processing, Tianjin, China, 17–19 October 2009; pp. 1–5.
69. Biswas, B.; Bhattacharya, U.; Chaudhuri, B.B. A Global-to-Local Approach to Binarization of Degraded Document Images. In Proceedings of the 2014 22nd International Conference on Pattern Recognition, Stockholm, Sweden, 24–28 August 2014; pp. 3008–3013.
70. Annabestani, M.; Saadatmand-Tarzjan, M. A New Threshold Selection Method Based on Fuzzy Expert Systems for Separating Text from the Background of Document Images. *Iran. J. Sci. Technol. Trans. Electr. Eng.* **2018**, *43*, 219–231. [[CrossRef](#)]
71. Gatos, B.; Pratikakis, I.; Perantonis, S.J. Improved document image binarization by using a combination of multiple binarization techniques and adapted edge information. In Proceedings of the 2008 19th International Conference on Pattern Recognition, Tampa, FL, USA, 8–11 December 2008; pp. 1–4.

72. Rosenfeld, A.; de la Torre, P. Histogram concavity analysis as an aid in threshold selection. *IEEE Trans. Syst. Man Cybern.* **1983**, SMC-13, 231–235.
73. Sezan, M.I. A Peak Detection Algorithm and its Application to Histogram-Based Image Data Reduction. *Comput. Vis. Graph. Image Process.* **1990**, 49, 36–51. [\[CrossRef\]](#)
74. Pavlidis, T. Threshold selection using second derivatives of the gray scale image. In Proceedings of the 2nd International Conference on Document Analysis and Recognition (ICDAR '93), Tsukuba Science City, Japan, 20–22 October 1993; pp. 274–277.
75. Kapur, J.N.; Sahoo, P.K.; Wong, A.K.C. A new method for gray-level picture thresholding using the entropy of the histogram. *Comput. Vis. Graph. Image Process.* **1985**, 29, 273–285. [\[CrossRef\]](#)
76. Abutableb, A.S. Automatic thresholding of gray-level pictures using two-dimensional entropy. *Comput. Vis. Graph. Image Process.* **1989**, 47, 22–32. [\[CrossRef\]](#)
77. Hertz, L.; Schafer, R.W. Multilevel thresholding using edge matching. *Comput. Vis. Graph. Image Process.* **1988**, 44, 279–295. [\[CrossRef\]](#)
78. Badekas, E.; Papamarkos, N. Optimal combination of document binarization techniques using a self-organizing map neural network. *Eng. Appl. Artif. Intell.* **2007**, 20, 11–24. [\[CrossRef\]](#)
79. Su, B.; Lu, S.; Tan, C.L. Combination of Document Image Binarization Techniques. In Proceedings of the 2011 International Conference on Document Analysis and Recognition, Beijing, China, 18–21 September 2011; pp. 22–26.
80. Pastor-Pellicer, J.; Boquera, S.E.; Zamora-Martínez, F.; Afzal, M.Z.; Bleda, M.J.C. Insights on the Use of Convolutional Neural Networks for Document Image Binarization. In Proceedings of the International Work-Conference on Artificial and Natural Neural Networks, Palma de Mallorca, Spain, 10–12 June 2015.
81. Pratikakis, I.; Gatos, B.; Ntirogiannis, K. H-DIBCO 2010—Handwritten Document Image Binarization Competition. In Proceedings of the 2010 12th International Conference on Frontiers in Handwriting Recognition, Kolkata, India, 16–18 November 2010; pp. 727–732.
82. Pratikakis, I.; Gatos, B.; Ntirogiannis, K. ICFHR 2012 Competition on Handwritten Document Image Binarization (H-DIBCO 2012). In Proceedings of the 2012 International Conference on Frontiers in Handwriting Recognition, Bari, Italy, 18–20 September 2012; pp. 817–822.
83. Fischer, A.; Indermühle, E.; Bunke, H.; Viehhauser, G.; Stolz, M. Ground truth creation for handwriting recognition in historical documents. In Proceedings of the International Workshop on Document Analysis Systems, Boston, MA, USA, 9–11 June 2010.
84. Fischer, A.; Frinken, V.; Fornés, A.; Bunke, H. Transcription alignment of Latin manuscripts using hidden Markov models. In Proceedings of the 2011 Workshop on Historical Document Imaging and Processing, Beijing, China, 16–17 September 2011.
85. Saddami, K.; Munadi, K.; Arnia, F. Degradation Classification on Ancient Document Image Based on Deep Neural Networks. In Proceedings of the 2020 3rd International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia, 24–25 November 2020; pp. 405–410.
86. Sandler, M.; Howard, A.G.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
87. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6848–6856.
88. He, S.; Schomaker, L. DeepOtsu: Document Enhancement and Binarization using Iterative Deep Learning. *Pattern Recognit.* **2019**, 91, 379–390. [\[CrossRef\]](#)
89. Vo, Q.N.; Kim, S.; Yang, H.J.; Lee, G. Binarization of degraded document images based on hierarchical deep supervised network. *Pattern Recognit.* **2018**, 74, 568–586. [\[CrossRef\]](#)
90. Meng, G.; Yuan, K.; Wu, Y.; Xiang, S.; Pan, C. Deep Networks for Degraded Document Image Binarization through Pyramid Reconstruction. In Proceedings of the 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, Japan, 9–15 November 2017; Volume 01, pp. 727–732.
91. Shelhamer, E.; Long, J.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
92. Tensmeyer, C.; Martinez, T.R. Document Image Binarization with Fully Convolutional Neural Networks. In Proceedings of the 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, Japan, 9–15 November 2017; Volume 01, pp. 99–104.
93. Ayyalasomayajula, K.R.; Malmberg, F.; Brun, A. PDNet: Semantic Segmentation integrated with a Primal-Dual Network for Document binarization. *Pattern Recognit. Lett.* **2018**, 121, 52–60. [\[CrossRef\]](#)
94. Riegler, G.; Ferstl, D.; Rüther, M.; Bischof, H. A Deep Primal-Dual Network for Guided Depth Super-Resolution. *arXiv* **2016**, arXiv:1607.08569.
95. Dumpala, V.; Kurupathi, S.R.; Bukhari, S.S.; Dengel, A.R. Removal of Historical Document Degradations using Conditional GANs. In Proceedings of the International Conference on Pattern Recognition Applications and Methods, Prague, Czech Republic, 19–21 February 2019.

96. Wang, T.C.; Liu, M.Y.; Zhu, J.Y.; Tao, A.; Kautz, J.; Catanzaro, B. High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8798–8807.
97. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–19 October 2017; pp. 2242–2251.
98. Kim, T.; Cha, M.; Kim, H.; Lee, J.K.; Kim, J. Learning to Discover Cross-Domain Relations with Generative Adversarial Networks. *Int. Conf. Mach. Learn.* **2017**, *70*, 1857–1865.
99. Suh, S.; Kim, J.; Lukowicz, P.; Lee, Y.O. Two-Stage Generative Adversarial Networks for Document Image Binarization with Color Noise and Background Removal. *arXiv* **2020**, arXiv:2010.10103.
100. Bhunia, A.K.; Bhunia, A.K.; Sain, A.; Roy, P.P. Improving Document Binarization Via Adversarial Noise-Texture Augmentation. In Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 22–25 September 2019; pp. 2721–2725.
101. Bhowmik, S.; Sarkar, R.; Das, B.; Doermann, D.S. GiB: A Game Theory Inspired Binarization Technique for Degraded Document Images. *IEEE Trans. Image Process.* **2019**, *28*, 1443–1455. [[CrossRef](#)] [[PubMed](#)]
102. Kumar, A.; Ghose, S.; Chowdhury, P.N.; Roy, P.P.; Pal, U. UDBNET: Unsupervised Document Binarization Network via Adversarial Game. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 7817–7824.
103. Konwer, A.; Bhunia, A.K.; Bhowmick, A.; Bhunia, A.K.; Banerjee, P.; Roy, P.P.; Pal, U. Staff line Removal using Generative Adversarial Networks. In Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018; pp. 1103–1108.
104. Zhao, J.; Shi, C.; Jia, F.; Wang, Y.; Xiao, B. Document image binarization with cascaded generators of conditional generative adversarial networks. *Pattern Recognit.* **2019**, *96*, 106968. [[CrossRef](#)]
105. Mirza, M.; Osindero, S. Conditional Generative Adversarial Nets. *arXiv* **2014**, arXiv:1411.1784.
106. Souibgui, M.A.; Kessentini, Y. DE-GAN: A Conditional Generative Adversarial Network for Document Enhancement. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 1180–1191. [[CrossRef](#)] [[PubMed](#)]
107. De, R.; Chakraborty, A.; Sarkar, R. Document Image Binarization Using Dual Discriminator Generative Adversarial Networks. *IEEE Signal Process. Lett.* **2020**, *27*, 1090–1094. [[CrossRef](#)]
108. Rajesh, B.; Agrawal, M.; Bhuva, M.; Kishore, K.; Javed, M. Document Image Binarization in JPEG Compressed Domain using Dual Discriminator Generative Adversarial Networks. In *Computer Vision and Machine Intelligence: Proceedings of CVMI 2022*; Springer: Singapore, 2022.
109. Lin, Y.S.; Ju, R.; Chen, C.C.; Lin, T.Y.; Chiang, J.S. Three-stage binarization of color document images based on discrete wavelet transform and generative adversarial networks. *arXiv* **2022**, arXiv:2211.16098.
110. Fathallah, A.; El-Yacoubi, M.A.; Amara, N.E.B. EHDI: Enhancement of Historical Document Images via Generative Adversarial Network. In Proceedings of the 18th International Conference on Computer Vision Theory and Applications, Lisbon, Portugal, 19–21 February 2023.
111. Guo, Y.; Ji, C.; Zheng, X.; Wang, Q.; Luo, X. Multi-scale Multi-attention Network for Moiré Document Image Binarization. *Signal Process. Image Commun.* **2021**, *90*, 116046. [[CrossRef](#)]
112. Peng, X.; Wang, C.; Cao, H. Document Binarization via Multi-resolutional Attention Model with DRD Loss. In Proceedings of the 2019 International Conference on Document Analysis and Recognition (ICDAR), Sydney, Australia, 20–25 September 2019; pp. 45–50.
113. Bezmaternykh, P.V.; Ilin, D.; Nikolaev, D.P. U-Net-bin: hacking the document image binarization contest. *Comput. Opt.* **2019**, *43*, 825–882. [[CrossRef](#)]
114. Zhao, P.; Wang, W.; Zhang, G.; Lu, Y. Alleviating pseudo-touching in attention U-Net-based binarization approach for the historical Tibetan document images. *Neural Comput. Appl.* **2021**, *35*, 13791–13802. [[CrossRef](#)]
115. Ma, K.; Shu, Z.; Bai, X.; Wang, J.; Samaras, D. DocUNet: Document Image Unwarping via a Stacked U-Net. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4700–4709.
116. Peng, X.; Cao, H.; Natarajan, P. Using Convolutional Encoder-Decoder for Document Image Binarization. In Proceedings of the 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, Japan, 9–15 November 2017; Volume 1, pp. 708–713.
117. Souibgui, M.A.; Biswas, S.; Jemni, S.K.; Kessentini, Y.; Forn’es, A.; Llad’os, J.; Pal, U. DocEnTr: An End-to-End Document Image Enhancement Transformer. In Proceedings of the 2022 26th International Conference on Pattern Recognition (ICPR), Montreal, QC, Canada, 21–25 August 2022; pp. 1699–1705.
118. Chaurasia, A.; Culurciello, E. LinkNet: Exploiting encoder representations for efficient semantic segmentation. In Proceedings of the 2017 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL, USA, 10–13 December 2017; pp. 1–4.
119. Xiong, W.; Jia, X.; Yang, D.; Ai, M.; Li, L.; Wang, S. DP-LinkNet: A convolutional network for historical document image binarization. *KSII Trans. Internet Inf. Syst.* **2021**, *15*, 1778–1797.

120. Zhou, L.; Zhang, C.; Wu, M. D-LinkNet: LinkNet with Pretrained Encoder and Dilated Convolution for High Resolution Satellite Imagery Road Extraction. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; pp. 192–1924.
121. Trier, Ø.D.; Taxt, T. Evaluation of Binarization Methods for Document Images. *IEEE Trans. Pattern Anal. Mach. Intell.* **1995**, *17*, 312–315. [\[CrossRef\]](#)
122. Ramírez-Ortegón, M.A.; Tapia, E.; Ramírez-Ramírez, L.L.; Rojas, R.; Jiménez, E.V.C. Transition pixel: A concept for binarization based on edge detection and gray-intensity histograms. *Pattern Recognit.* **2010**, *43*, 1233–1243. [\[CrossRef\]](#)
123. Ntirogiannis, K.; Gatos, B.; Pratikakis, I. Performance Evaluation Methodology for Historical Document Image Binarization. *IEEE Trans. Image Process.* **2013**, *22*, 595–609. [\[CrossRef\]](#)
124. Ntirogiannis, K.; Gatos, B.; Pratikakis, I. ICFHR2014 Competition on Handwritten Document Image Binarization (H-DIBCO 2014). In Proceedings of the 2014 14th International Conference on Frontiers in Handwriting Recognition, Crete Island, Greece, 1–4 September 2014; pp. 809–813.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.