Preprints.org

Review

Summary of Document Image Binarization

Zhengxian Yang , Rui Zhang , Yanxi Zhou , Jinlong He , Jianwen Shi , Shikai Zuo [*]

Review

Summary of Document Image Binarization

Zhengxian Yang, Rui Zhang, Yanxi Zhou, Jinlong He, JianWen Shi and Shikai Zuo *

Xiamen University of Technology, Fujian, China

* 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 degree of document image, researchers have proposed a variety of solutions. This paper reviews the main binarization techniques, including both traditional algorithms and deep learning-based algorithms. We also summarize some difficulties and challenges in the field of document image binarization. Here, we evaluate various image binarization techniques to identify shortcomings in current methods and provide some help for future research.

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

---

1. Introduction

Image binarization is an important aspect of image analysis, such as scene text detection and medical image analysis. 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. 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. In the 1960s, research on threshold segmentation primarily focused on global thresholding, local thresholding, and adaptive thresholding methods. 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 (CNN), generative adversarial networks (GAN), and recurrent neural networks (RNN), in the field of document image binarization. Despite the extensive research in digital image processing, there remain numerous unresolved challenges in dealing with degraded document images. The continuous emergence of new algorithms and technologies offers opportunities for further optimization in the field of document image binarization.

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. Examles of degraded document images from DIBCO.

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 [1] 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 can be expressed as follows:

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

We represent the image pixel in the gray level of the image, the image has L-order gray level, $\omega_0\left(T\right)$ and $\omega_1\left(T\right)$ are the probability distribution of the target and background when the threshold value is T, $\mu_0\left(T\right)$ and $\mu_1\left(T\right)$ 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, otherwise it is black.

The Otus 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. Consequently, there is no universally effective thresholding method for handling such contaminated images.

2.2. Local threshold method

The Niblack [2] 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:

$$T = m + k * s \qquad (2)$$

Where, m represents the average gray value of pixels in the local area, s represents the standard deviation, and k is a constant, a correction factor, which can be adjusted according to the foreground and background conditions of the image.

Trier [3] believes that Niblack performs better than other local binarization methods in gray images with low contrast, noise and uneven background intensity. However, Saxena [4] proposed that the window size is the main defect of the local threshold method. Both large and small size Windows will generate noise, and even in Windows without target pixels, the local threshold method can still detect target pixels. Furthermore, the time taken to compute the threshold is proportional to the square of the window size. Similarly, Chaki [5] asserts 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 becomes challenging. Even with an accurate k value, it still generates pepper noise in the shadowed areas of the image or in non-text regions.

The Sauvola [6] 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:

$$T(x,y) = m(x,y) * \left[1 + k * \left(\frac{s(x,y)}{R} - 1 \right) \right] \qquad (3)$$

It can be seen from the formula that Sauvola introduces a new parameter R, which represents the dynamic range of standard deviation. However, it still requires the use of an artificial k value and window size for threshold calculation. In cases of low contrast targets in a document, such as textured backgrounds or translucent false images, the Sauvola algorithm may either remove or partially recover them. 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 [7].

Wolf [8] and colleagues conducted a global statistical normalization of image contrast and average gray value. This enabled the automatic detection of text regions and adaptive threshold selection based on the characteristics of the text regions. This process optimizes the binarization of low-contrast images using Sauvola's algorithm. In a similar vein, Gatos [9] and his team estimated the background surface of documents using the binary document image generated by Sauvola's thresholding algorithm. This method eliminates the need for manual parameter adjustment and effectively addresses issues related to degraded documents, such as shadows, non-uniform illumination, and low contrast. Mustafa et al. [10] proposed "WAN" algorithm to improve the lost detail strokes by raising the threshold of binarization based on Sauvola algorithm.

In addressing the issue of black noise in the Niblack algorithm, Khurshid et al. [11] introduced the NICK algorithm, 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:

$$T = m + k\sqrt{\frac{\sum p_i^2 - m^2}{NP}} \tag{4}$$

Pi 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, although this may lead to interrupted characters or faint drawings. Conversely, when the k value is close to 0.1, the text can be extracted with complete clarity, but some noise is retained. However, the selection of the k value requires manual adjustment if the characters are thin or if the text document has low contrast. Consequently, B. Bataineh [12] contends that the method does not outperform the Niblack algorithm in exceptional circumstances, such as very low-contrast images, variations in text size and thickness, or finer characters with low contrast.

Su et al. (2010) [13] introduced a novel image contrast technique that utilizes local image maxima and minima instead of the image gradient. This technique is particularly beneficial for unevenly lit and degraded documentation. The method involves constructing a contrast image, detecting high contrast areas near the boundary of the character pencil, and using local thresholds to segment the text documentation. In a subsequent study in 2013, Su et al. [14] proposed a technique that combines local image contrast and gradation. This involved constructing an adaptive contrast map for the degraded document image input, dualizing the contrast image, and combining it with the Canny edge image to identify the edges of the text pencil image. Subsequently, the text paper was partitioned using local thresholds, with the threshold value serving as an estimate of the intensity of the border depicting the detected text within the window door. This method was employed to obtain high-quality data on Bickley paper.

In the local binary method, the window size is a crucial factor that can be customized by the user. 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. Therefore, custom settings may not be universally applicable to all images. Bataineh et al. [12] proposed a threshold approach based on dynamic flexible windows to address the issue of low contrast between the foreground and background, as well as the text in pencil drawings. This method involves two approaches: dynamically segmenting images into windows based on image characteristics and determining the appropriate threshold for each window. The computation time for the local average generally depends on the window size. In contrast, T. Romen Singh et al. [15] 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.

The global thresholding method is useful for document images that are obviously intermediate between foreground and background, or for images distinct bimodal histograms because it segments the foreground and background using a single threshold. However, this method is less robust and is not suited to images with low contrast or uneven illumination, hence it is usually employed for simple document photographs with a uniformly pure background. The local thresholding method uses several different thresholds to segment the image, which can be used for multi-target segmentation, but the disadvantage is that the segmented targets are poorly connected and usually bring about the phenomenon of ghosting, i.e., pseudo-strokes appearing in the background region, and the binarization result is greatly affected by the noise. In addition, some local thresholding approaches can bring character stroke breakage to some camera photos. We summarize the global threshold and local threshold methods in Table 1.

Table 1. Traditional binarization techniques(1).

Classification	Algorithm	Description
Global Threshold	Otsu [1]	The gray level corresponding to the maximum inter-class variance is selected as the global threshold.
Local Threshold	Niblack [2]	Calculate the mean and standard deviation of each pixel in the local window as the threshold value.
	Sauvola [6]	Introducing the dynamic range of standard deviation based on Niblack[49].
	Wolf [8]	Global statistical normalization based on Sauvola for automatic detection of text regions.
	Gatos [9]	Local adaptive based approach to enhance degraded image binarization documents.
	WAN [10]	Improving lost detail strokes by increasing the binarization threshold based on Sauvola.
	NICK [11]	In order to solve the black noise problem in Niblack, an improved method of threshold calculations was used.
	Su [13]	Image contrast technique defined using local image maxima and minima.
	Su [14]	Combination method based on local image contrast and local image gradient.
	Bataineh [12]	Dynamic segmentation of the image into windows based on image features, determining the threshold value for each window.
	T.R.Singh [15]	Local adaptive threshold segmentation uses the local mean and mean difference to remove the background.

* Global threshold and local threshold techniques.

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. [16] integrated Otsu and Bernsen's method. Zemouri et al. [17] enhanced the document picture using global thresholding before binarization and then applied a local thresholding strategy for binarization. Chaudhary et al. [18] 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. [19] 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 [20] developed a hybrid thresholding technique and determined the trade-off between local and global content using variational optimization. Xiao et al. [21] 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. [22] 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. [23] suggested a classification system to deal with degraded document photographs by blending the modification of local and global binarization algorithms.

He [24] compared Niblack, Sauvola and their adaptive threshold method in the article, and found that adaptive Niblack and adaptive Sauvola performed slightly better than originals. Adaptive thresholding scans the image with a sliding window centered on a pixel and compares the centroid pixel with pixels in its neighborhood to obtain different thresholds for different pixels. Usually, the fixed threshold is manually set according to the specific situation of different tasks, whereas the adaptive thresholding tends to estimate the background surface of the document first, and then the thresholds are calculated according to the estimated background surface. Bernsen's algorithm [25] calculates a separate threshold for each pixel based on the neighborhood of the pixel, which is the classical adaptive thresholding method. Moghaddam et al. [26] estimate the backdrop surface of the document by an adaptive and iterative image averaging approach. Messaoud et al. [27] apply a binarization technique to selected items of interest by combining a preprocessing stage and a localization step. Pardhi et al. [28] construct local thresholds by a combination of local image contrast and gradient combination to segment text and it also an adaptive image contrast technique. Kligler et al. [29] introduced a novel and generalized algorithm to replace the grayscale map as input to the algorithm with a visible-based low-light map, claiming that by doing so the values of the text pixels are closer to each other and better separated from those of the non-text pixels, and that the performance of the binary algorithm without changing the rest of the algorithm will also be Improvement. Adaptive approaches are generally able to handle part of the complexity of document images, but they often overlook the edge aspects of image features and can lead to artifacts. We summarize these methods in the Table 2.

Table 2. Traditional binarization techniques(2).

Classification	Algorithm	Description
Mixed threshold method	Yang [16]	Combining the methods of Otsu and Bernsen.
	Zemouri [17]	Combination of global and local thresholds.
	Chaudhary [18]	Estimating the background first before using the hybrid thresholding algorithm.
	K.Ntirogiannis [19]	Global and local adaptive binarization.
	Liang [20]	Hybrid thresholding approach with variational optimization method to determine the trade-off between local and global content.
	Xiao [21]	Binarization is achieved by a global branch and a local branch.
	Saddami [22]	Combined local and global threshold processing methods.
	Ranjitha [23]	A degraded document image classification system based on mixed threshold is proposed.
Adaptive method	Bernsen [25]	Compute a separate threshold for each pixel based on the Adaptive pixel's neighborhood.
	Moghaddam [26]	Estimation of the background surface of a document by an adaptive and iterative image-averaging procedure.
	Messaoud [27]	Combination of pre-processing step and positioning step.
	Pardhi [28]	A combination of local image contrast and gradient combination to segment text.
	Kligler [29]	Replace the grayscale map with a visible low-light map as the input to the algorithm.

*Mixed threshold method adaptive method.

2.4. Image feature method

Based on the threshold method of edge detection, we first identifying the edge pixels within the image and then using 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 the images. 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 the practical application. Santhanaprabhu et al. [30] applied the Sobel edge detection technique to extract text and perform document image binarization. Lu et al. [31] used the L1-norm image gradient to identify the edge of the font from the compensated document image. T. Lelore et al. [32] have described a quick solution for repairing document images by employing the edge-based FAIR 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. [33] exploited adaptive image contrast in combination with Canny's edge diagram to identify the edge pixels of the font. Jia et al. [34] used structural symmetry pixels (SSPs) to calculate local thresholds for the neighborhood. SSP is defined as the pixel around the stroke, whose gradient size is large enough and the direction is symmetric and opposite. The author extracts SSP 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 SSP detection. Hadjadj et al. [35] 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.

In general, these methods have a good effect in the processing of degraded printed document images or handwritten text images, compared to the simple thresholding method, the edge detection-based separation method can effectively extract text contour information, the detection speed is fast, but because the method is based on the calculation of the pixel gradient of the image itself, it is more sensitive to light changes and susceptible to the interference of light changes.

Fuzzy theory defines a fuzzy set, calculates the membership degree of each pixel belonging to each set through fuzzy logic operations. Finally, it classifies pixels based on their membership degree, to achieve the purpose of segmentation. The idea of clustering is to find the correlation from a set of unlabeled data to find the similarity between the data. Image segmentation itself can also be regarded as a clustering process. A. Lai et al. [36] used the K-means clustering algorithm to binarize document images. Tong et al. [37] combined the Niblack algorithm and the FCM algorithm 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. Soua et al. [38] proposed the K-means method (HBK) based on hybrid binarization and implemented real-time processing of the parallel HBK method in an OCR system. Among these methods, although K-means is widely used, due to its hard clustering nature and sensitivity to noise, some researchers have proposed more flexible soft clustering algorithms, such algorithm is the fuzzy C-means clustering algorithm (FCM), which can accommodate uncertainties related to data points. There are also clustering algorithms such as PFCM [39] and KFCM [40]. In addition, there is the FuzBin-based binarization method, which Annabestani et al. [41] use to extract text information from document images. They enhance image contrast with FESs and then combine FES with a pixel counting algorithm to obtain a range of threshold values. The middle value is taken as the final threshold. A method based on mathematical morphology operations to enhance fuzzy stroke information [9]. Biswas et al. [42] used a Gaussian filter to blur the input degraded image file. The application of Support Vector Machine (SVM) in processing historical document images. Xiong et al. [43] used SVM to classify image blocks into different categories based on statistical information such as mean, variance, and histogram of regions. They also determined the optimal global threshold for a preliminary segmentation of the foreground and background. After the image is segmented, the stroke width is estimated using the progressive scanning method. However, this method only considers a single class method, and the accuracy of the recognition results is not ideal.

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 [44–46]; entropy-based methods such as [47]; space binarization-based methods such as [48]; and object property-based methods such as [49], 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. We summarize the above methods in the Table 3.

Table 3. Traditional binarization techniques(3).

Classification	Algorithm	Description
Edge detection	Santhanaprabhu [30]	Text extraction and document image binarization using Sobel edge detection.
	Lu [31]	Combining L1 parametric image gradient and local threshold segmentation.
	T.Lelore [32]	FAIR algorithm based on edge detection to locate text.
	Holambe [33]	Adaptive image contrast combined with Canny's edge map to Edge detection identify stroke-edge pixels.
	Jia [34]	Use structural symmetric pixels (SSPs) to calculate local thresholds in the neighborhood.
Fuzzy Logic	A.Lai [36]	Binarization of document images with K-means clustering algorithm.
	Tong [37]	Niblack and fuzzy C-Means algorithm (FCM) were used to Clustering calculate the local thresholds, respectively.
	soua [38]	a K-means method based on hybrid binarization and implemented real-time processing of the parallel HBK method in an OCR system.
	Annabestani [41]	A global threshold selection method based on fuzzy expert systems (FESs).
	Biswas [42]	Blurring of the input degraded file image with Gaussian filter.

* Binarization method based on image features.

3. Deep learning binarization techniques

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. 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 CNN

　　Convolutional Neural Network (CNN) is 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 the CNN. In the field of document image binarization, a CNN 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 CNN with other techniques, such as regularization and data enhancement, to optimize performance and enhance the model's generalization ability. In their paper, Pastor-Pellicer et al. [50] describe the practical application of CNN in the task of document image binarization. They also verify the effectiveness of CNN using image datasets of handwritten and printed documents. Saddami et al. [51] compared the degradation classification of three pre-trained CNN architecture models, namely Resnet101 [52], Mobilenet V2 [53], and Shufflenet [54], in ancient document images. Shufflenet achieved superior performance in terms of accuracy and computational efficiency. He et al. [55] 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. Vo et al. [56] 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. [57] 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. Badekas et al. [58] propose an integrated system for binarizing normal and degraded printed documents to enhance the visualization and recognition of textual characters. They learn from the binarization results produced by various techniques using neural networks. While this approach is highly effective for files with complex backgrounds and images, it can also result in lengthy processing times [59].

　　In 2015, Long et al. [60] proposed Fully Convolutional Networks (FCN), 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 [61] 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 not sufficiently sensitive to the details in image processing, and the results obtained are not refined enough. In tasks that require high levels of detail, such as image processing of ancient document image, FCN has potential for further improvement. Ayyalasomayajula et al. [62] proposed an end-to-end structure that combines the FCN and the Primal-Dual network (PD-Net [63]) 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 the Table 4.

Table 4. Deep learning techniques(1).

Algorithm	Description
Beldekas [58]	Learning from binarization results generated by different techniques using neural networks.
Pastor-Pellicer [50]	Describes a practical application of CNNs in document image binarization tasks.
Meng [57]	Deep Convolutional Neural Network (DCNN) based framework.
Vo [56]	A new supervised binarization method based on deep supervised networks (DSN).
He [55]	Combining the traditional Otsu algorithm with a convolutional neural network, called DeepOtsu.
Saddami [51]	Three pre-trained CNN architectures were compared, namely Resnet101, Mobilenet V2 and Shufflenet.
Tensmeyer [61]	FCN-based binarization algorithm for low-quality document images and palm leaf manuscript images.
Ayyalasoma-yajula [62]	Combining a Fully Convolutional Neural Network (FCN) and a Primal-Dual Network (PD-Net).

* Based on CNN.

3.2. Based on GANs

Generative Adversarial Networks (GANs) [64] are composed of generator networks and discriminator networks. GAN performs well in binarization, text region detection, and text recognition, among other applications. It has a wide range of applications in image processing, as mentioned in [65–68]. GAN transforms 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. [69] proposed a two-stage GANs 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. [70] 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. Kumar et al. [71] optimized Bhunia's algorithm by introducing a joint discriminator to combine TANet and unsupervised document binarization network (UDBNet) to resolve dataset bias and achieve better performance on actual degraded images. Konwer et al. [72] used GAN to remove staff line to achieve binarization in the pre-processing step of optical music recognition. Zhao et al. [73] introduced conditional generation adversarial networks (cGANs) [74] to solve the problem of multi-scale information composition in binary tasks. Souibgui et al. [75] used cGANs to propose a pix2pix framework called DE-GAN (Document Enhancement Generative Adversarial Network)

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. As a result, this method can also effectively remove watermarks or stamps. The author validates the model's accuracy using multiple public data sets. R. De et al. [76] 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 is used to solve the issue of class imbalance among pixels. Rajesh et al. [77] 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. [78] 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 the Table 5.

Table 5. Deep learning techniques(2).

Algorithm	Description
Bhunia [70]	A texture enhancement network is constructed by introducing an adversarial learning approach.
Zhao [73]	cGANs are introduced to solve the multi-scale information combination problem in binarization tasks.
Kumar [71]	Improving the algorithm of Bhunia, the joint discriminator combines the ATANet and the UDBNet.
Suh [69]	A two-stage color document image enhancement and binarization method using GANs.
Souibgui [75]	A DE-GAN framework is proposed to restore severely degraded document images.
R.De [76]	Lossy DD-GANs using focal loss as a generator.
Rajesh [77]	DD-GANs are applied, proposing direct binarization using JPEG compressed document images.
Fathallah [79]	A historical document image enhancement model is proposed based on GANs.

* Based on GANs.

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. [80] proposed a novel Multi-scale Multi-attention Network (MsMa-Net) for the fresh moiré document image binarization task. Peng et al. [81] 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 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. [82] used the U-Net architecture to propose a CNN-based method called U-Net-bin, which won the first place in the DIBCO '17 competition. Xiao et al. [21] 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. [83] 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. [84] 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. [85] 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. [86] 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. [87] proposed a network architecture called LinkNet, which drew inspiration from the U-Net model and adopted Encoder-Decoder structure to create a lightweight network capable of real-time segmentation. Xiong et al. [88] proposed an improved semantic segmentation model called DP-LinkNet, which is based on the LinkNet and D-LinkNet models [89]. 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.

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 attention mechanism in the Table 6.

Table 6. Deep learning techniques(3).

Algorithm	Description
Peng [85]	The probabilities of text regions are inferred from a multi-resolution attention model, which is then fed into a ConvCRF.
Ke Ma [84]	A hybrid model based on U-Net and Transformer.
Peng [81]	Convolutional encoder-decoder model with deep learning.
Bezmaternykh [82]	It uses UNet architecture to achieve more accurate historical document binarization.
Xiao [21]	A local-global combination method based on U-Net.
Guo [80]	A Multiscale Multi-Attention Network (MsMa-Net).
Zhao [83]	A U-Net based binarization method for Tibetan historical document images.
Xiong [88]	Adopts the D-LinkNet architecture as its backbone, with the proposed HDC and SPP modules between the encoder and the decoder.

* Based on attention mechanism.

4. Results

This review covers literature on document image binarization, ranging from traditional binarization algorithms to binarization based on deep learning models. Above, we have summarized the methods used in relevant literature to provide a reference for this research direction. Next, we will present some commonly used indicators for evaluating the quality of document image binarization. Additionally, we will provide experimental results based on the H-DIBCO2016 dataset.

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, and 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.

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

$$MSE = \frac{\sum\limits_{x=1}^{M} \sum\limits_{y=1}^{N} \left(I_{bin}\left(x,y\right) - I_{GT}\left(x,y\right) \right)^2}{MN} \tag{6}$$

Where MAXI denotes the maximum image pixel value, usually 255, 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 is a metric to evaluate the binarization results, which combines Precision and Recall, and is the harmonic mean of the two value. The following equation can express it:

$$F - Measure = \frac{\left(1 + \beta^2\right) Precision * Recall}{\left(\beta2 * Precision + Recall\right)} \tag{7}$$

Where β is a weighting factor, generally taking a value of 1, indicating that Precision and Recall are equally important. When β=1, also called F1-Score, Precision indicates the proportion of pixels that are binarized as foreground that really belong to the foreground, while Recall indicates the proportion of pixels that really belong to the foreground that is correctly binarized as foreground. The two are contradictory metrics, and in the task of document image binarization, we need to minimize the loss of text information due to binarization errors, so we need a higher Precision. The two are calculated as follows:

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

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

Where TP (True Positive) indicates True Positive, the number of pixels correctly classified as foreground; FP (False Positive) indicates False Positive, the number of pixels incorrectly classified as foreground; and FN (False Negative) indicates False Negative, the number of pixels incorrectly classified as background.

4.1.3. Pseudo F-measure

Pseudo F-measure $\left(F_{ps}\right)$, an improved algorithm for F-measure, is mainly used for the evaluation of binarization, which is calculated as:

$$F_{ps} = \frac{\left(1 + \beta^2\right) * Precision * pRecall}{\left(\beta2 * Precision + pRecall\right)} \tag{10}$$

Differs from F-measure in that Pseudo F-measure introduces the pseudo-recall (pRecall) [90] concept. In [91], the pseudo-recall rate refers to the percentage of binarized baseline images after refinement processing.

4.1.4. DRD

Distance Reciprocal Distortion(DRD), is a metric for image quality evaluation (cf. [92]), mainly used to measure the sharpness and contrast of the image. DRD is used to measure the visual distortion in binary document images.

$$DRD = \frac{\sum\limits_{k=1}^{N} DRD_k}{NUBN} \tag{11}$$

Where DRD_k denotes the distortion of the kth flipped pixel, and NUBN denotes the number of non-uniform color blocks in the reference image (GT).

4.2. Experimental result

The experimental data in this paper is based on the dataset H-DIBCO2016 [93] of the Handwritten Document Image Binarization Competition, which includes 10 degraded historical handwritten document images and their corresponding baseline images (Ground Truth, GT).

The Table 7 displays the evaluation results of the document image binarization method. All the performance of deep learning models-based methods is higher than that of traditional threshold-based binarization methods. According to the results, Kumar et al. [71] 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. [78]. Among threshold-based methods, the Otsu algorithm [1], Wolf algorithm [8], and Gatos algorithm [9] 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 [55], Zhao [73], and Peng [81] have shown promising results.

Table 7. Performance measures.

Classification	Algorithm	F-Measure↑	$F_p s$ ↑	PSNR↑	DRD↓
Traditional based	Otsu [1]	86.59	89.92	17.79	6.89
	Niblack [2]	63.10	63.64	12.47	31.48
	Sauvola [6]	81.27	83.91	15.36	17.92
	Wolf [8]	86.36	90.46	17.11	9.01
	NICK [11]	80.86	82.65	15.15	19.24
	Su [14]	74.41	87.03	15.52	12.55
	T.R.Singh [15]	84.47	87.56	16.99	9.73
	Bataineh [12]	83.27	84.43	16.02	15.61
	Wan [10]	75.14	76.06	13.61	30.33
	Gatos [9]	86.59	89.09	17.47	7.01
Deep learning based	Vo [56]	90.10	93.57	19.01	3.58
	Tensmeyer [61]	89.52	93.76	18.67	3.76
	He [55]	91.4	94.3	19.60	2.9
	Meng [57]	89.90±4.55	-	18.79±3.36	-
	Ayyalasomayajula [62]	90.18	-	18.99	3.61
	Zhao [73]	91.66	94.58	19.64	2.82
	Kumar [71]	93.4	96.2	20.1	2.2
	Lin [78]	91.46	96.32	19.66	2.94
	R.De [76]	89.98	95.23	18.83	3.61
	Xiao [21]	90.77	94.21	19.33	3.11
	Peng [81]	91.68	-	19.59	2.93
	Peng [85]	88.07±4.86	-	18.13±3.13	-

* The deep learning-based data comes from their original authors.

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. In recent years, the emergence of deep learning has presented new opportunities and challenges in the study of document image binarization. These algorithms have demonstrated remarkable achievements in the field of document image processing. Therefore, this paper mainly summarizes a range of traditional binarization methods as well as methods based on deep learning models. Traditional binarization algorithms perform well in processing document images with simple backgrounds, but they have limited effectiveness on complex background images. This includes images with mixed noise and heavily polluted document images. Deep learning methods achieve good results by segmenting image pixels, while also being capable of handling intricate details.

The application scenarios for document image binarization are very broad. Due to the variations in characters, numbers, and symbols across different countries, enhancing the ability of cross-language and character processing has become a crucial area of research. It will be a future research direction to design a corresponding binarization algorithm for document images in different application scenarios. Since the data characteristics of degraded document images are extremely complex, each image may exhibit different types and levels of damage or degradation. Furthermore, the main challenge in current network training is the lack of ground truth, which leads to inadequate data sets. Therefore, future work can be improved in this area, for example, by adopting unsupervised learning methods. The efficiency and accuracy of document image binarization can be improved by leveraging knowledge from various fields, such as image processing and computer vision.

References

1. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66.
2. Niblack, W. An introduction to digital image processing. 1986.
3. Trier, Ø.D.; Jain, A.K. Goal-Directed Evaluation of Binarization Methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **1995**, *17*, 1191–1201.
4. Saxena, L.P. Niblack's binarization method and its modifications to real-time applications: a review. *Artificial Intelligence Review* **2019**, *51*, 673–705.
5. Chaki, N.; Shaikh, S.H.; Saeed, K. A Comprehensive Survey on Image Binarization Techniques. 2014.
6. Sauvola, J.J.; Seppänen, T.; Haapakoski, S.; Pietikäinen, M. Adaptive document binarization. *Proceedings of the Fourth International Conference on Document Analysis and Recognition* **1997**, *1*, 147–152 vol.1.
7. Lazzara, G.; Géraud, T. Efficient Multiscale Sauvola's Binarization. *International Journal of Document Analysis and Recognition (IJDAR)* **2014**, *17*, 105–123. doi:10.1007/s10032-013-0209-0.
8. Wolf, C.; Jolion, J.M. Extraction and recognition of artificial text in multimedia documents. *Formal Pattern Analysis & Applications* **2003**, *6*, 309–326.
9. Gatos, B.; Pratikakis, I.; Perantonis, S.J. Improved document image binarization by using a combination of multiple binarization techniques and adapted edge information. *2008 19th International Conference on Pattern Recognition* **2008**, pp. 1–4.
10. Mustafa, W.A.; Kader, M.M.M.A. Binarization of Document Image Using Optimum Threshold Modification. *Journal of Physics: Conference Series* **2018**, *1019*.

11. Khurshid, K.; Siddiqi, I.; Faure, C.; Vincent, N. Comparison of Niblack inspired binarization methods for ancient documents. *Electronic imaging*, 2009.

12. 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.

13. Su, B.; Lu, S.; Tan, C.L. Binarization of historical document images using the local maximum and minimum. International Workshop on Document Analysis Systems, 2010.

14. Su, B.; Lu, S.; Tan, C.L. Robust Document Image Binarization Technique for Degraded Document Images. *IEEE Transactions on Image Processing* **2013**, *22*, 1408–1417.

15. Singh, T.R.; Roy, S.; Singh, O.I.; Sinam, T.; Singh, K.M. A New Local Adaptive Thresholding Technique in Binarization. *ArXiv* **2012**, *abs/1201.5227*.

16. Yang, Y. OCR Oriented Binarization Method of Document Image. *2008 Congress on Image and Signal Processing* **2008**, *4*, 622–625.

17. Zemouri, E.T.; Chibani, Y.; Brik, Y. Enhancement of Historical Document Images by Combining Global and Local Binarization Technique. *International Journal of Information Engineering and Electronic Business* **2014**, *4*.

18. Chaudhary, P.; Ambedkar, B. AN EFFECTIVE AND ROBUST TECHNIQUE FOR THE BINARIZATION OF DEGRADED DOCUMENT IMAGES. *International Journal of Research in Engineering and Technology* **2014**, *03*, 140–145.

19. Ntirogiannis, K.; Gatos, B.; Pratikakis, I. A combined approach for the binarization of handwritten document images. *Pattern Recognit. Lett.* **2014**, *35*, 3–15.

20. Liang, Y.; Lin, Z.; Sun, L.; Cao, J. Document image binarization via optimized hybrid thresholding. *2017 IEEE International Symposium on Circuits and Systems (ISCAS)* **2017**, pp. 1–4.

21. Xiao, H.; Lin, L.; Rong, L.; Chengshen, X.; Ye, M. Binarization of degraded document images with global-local U-Nets. *Optik* **2020**, *203*, 164025.

22. Saddami, K.; Arnia, F.; Away, Y.; Munadi, K. Kombinasi Metode Nilai Ambang Lokal dan Global untuk Restorasi Dokumen Jawi Kuno. 2020.

23. Ranjitha, P.; Shreelakshmi, T.D. A Hybrid Ostu based Niblack Binarization for Degraded Image Documents. *2021 2nd International Conference for Emerging Technology (INCET)* **2021**, pp. 1–7.

24. He, J.; Do, Q.; Downton, A.C.; Kim, J.H. A comparison of binarization methods for historical archive documents. *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)* **2005**, pp. 538–542 Vol. 1.

25. Bernsen, J. Dynamic thresholding of grey-level images **1986**.

26. Moghaddam, R.F.; Cheriet, M. RSLDI: Restoration of single-sided low-quality document images. *Pattern Recognit.* **2009**, *42*, 3355–3364.

27. Messaoud, I.B.; Amiri, H.; Abed, H.E.; Märgner, V. New Binarization Approach Based on Text Block Extraction. *2011 International Conference on Document Analysis and Recognition* **2011**, pp. 1205–1209.

28. Pardhi, S.; Kharat, D.G.U. An Improved Binarization Method for Degraded Document. 2017.

29. Kligler, N.; Katz, S.; Tal, A. Document Enhancement Using Visibility Detection. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* **2018**, pp. 2374–2382.

30. Santhanaprabhu, G.; M., E.K.; Karthick, B.; M., E.K.; Srinivasan, P.; Vignesh, R.K.; Sureka, K. Extraction and Document Image Binarization Using Sobel Edge Detection. 2014.

31. Lu, S.; Su, B.; Tan, C.L. Document image binarization using background estimation and stroke edges. *International Journal on Document Analysis and Recognition (IJDAR)* **2010**, *13*, 303–314.

32. Lelore, T.; Bouchara, F. FAIR: A Fast Algorithm for Document Image Restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2013**, *35*, 2039–2048.

33. Holambe, S.N.; Shinde, U.B.; Choudhari, B.S. Image Binarization for Degraded Document Images. *International Journal of Computer Applications* **2015**, *128*, 38–43.

34. Jia, F.; Shi, C.; He, K.; Wang, C.; Xiao, B. Document Image Binarization Using Structural Symmetry of Strokes. *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)* **2016**, pp. 411–416.

35. Hadjadj, Z.; Cheriet, M.; Meziane, A.; Cherfa, Y. A new efficient binarization method: application to degraded historical document images. *Signal, Image and Video Processing* **2017**, *11*, 1155–1162.

36. Lai, A.N.; Lee, G. Binarization by Local K-means Clustering for Korean Text Extraction. *2008 IEEE International Symposium on Signal Processing and Information Technology* **2008**, pp. 117–122.

37. Tong, L.; Chen, K.; Zhang, Y.; Fu, X.L.; Duan, J. Document Image Binarization Based on NFCM. *2009 2nd International Congress on Image and Signal Processing* **2009**, pp. 1–5.

38. Soua, M.; Kachouri, R.; Akil, M. GPU parallel implementation of the new hybrid binarization based on Kmeans method (HBK). *Journal of Real-Time Image Processing* **2018**, *14*, 363–377.

39. Pal, N.R.; Pal, K.; Keller, J.M.; Bezdek, J.C. A possibilistic fuzzy c-means clustering algorithm. *IEEE Transactions on Fuzzy Systems* **2005**, *13*, 517–530.

40. Farahmand, A.; Sarrafzadeh, H.; Shanbehzadeh, J. Noise removal and binarization of scanned document images using clustering of features. 2017.

41. Annabestani, M.; Saadatmand-Tarzjan, M. A New Threshold Selection Method Based on Fuzzy Expert Systems for Separating Text from the Background of Document Images. *Iranian Journal of Science and Technology, Transactions of Electrical Engineering* **2018**, *43*, 219–231.

42. Biswas, B.; Bhattacharya, U.; Chaudhuri, B.B. A Global-to-Local Approach to Binarization of Degraded Document Images. *2014 22nd International Conference on Pattern Recognition* **2014**, pp. 3008–3013.

43. 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.

44. Rosenfeld, A.; de la Torre, P. Histogram concavity analysis as an aid in threshold selection. *IEEE Transactions on Systems, Man, and Cybernetics* **1983**, *SMC-13*, 231–235.

45. 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.

46. Pavlidis, T. Threshold selection using second derivatives of the gray scale image. *Proceedings of 2nd International Conference on Document Analysis and Recognition (ICDAR '93)* **1993**, pp. 274–277.

47. 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.

48. Abutableb, A.S. Automatic thresholding of gray-level pictures using two-dimensional entropy. *Graphical Models graphical Models and Image Processing computer Vision, Graphics, and Image Processing* **1989**, *47*, 22–32.

49. Hertz, L.; Schafer, R.W. Multilevel thresholding using edge matching. *Comput. Vis. Graph. Image Process.* **1988**, *44*, 279–295.

50. 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. International Work-Conference on Artificial and Natural Neural Networks, 2015.

51. Saddami, K.; Munadi, K.; Arnia, F. Degradation Classification on Ancient Document Image Based on Deep Neural Networks. *2020 3rd International Conference on Information and Communications Technology (ICOIACT)* **2020**, pp. 405–410.

52. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* **2015**, pp. 770–778.

53. Sandler, M.; Howard, A.G.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* **2018**, pp. 4510–4520.

54. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* **2017**, pp. 6848–6856.

55. He, S.; Schomaker, L. DeepOtsu: Document Enhancement and Binarization using Iterative Deep Learning. *Pattern Recognit.* **2019**, *91*, 379–390.

56. 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.

57. Meng, G.; Yuan, K.; Wu, Y.; Xiang, S.; Pan, C. Deep Networks for Degraded Document Image Binarization through Pyramid Reconstruction. *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)* **2017**, *01*, 727–732.

58. 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.

59. Su, B.; Lu, S.; Tan, C.L. Combination of Document Image Binarization Techniques. *2011 International Conference on Document Analysis and Recognition* **2011**, pp. 22–26.

60. Shelhamer, E.; Long, J.; Darrell, T. Fully convolutional networks for semantic segmentation. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* **2014**, pp. 3431–3440.

61. Tensmeyer, C.; Martinez, T.R. Document Image Binarization with Fully Convolutional Neural Networks. *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)* **2017**, *01*, 99–104.

62. 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.

63. Riegler, G.; Ferstl, D.; Rüther, M.; Bischof, H. A Deep Primal-Dual Network for Guided Depth Super-Resolution. *ArXiv* **2016**, *abs/1607.08569*.

64. Treat, I.; Yoon, J. GENERATIVE ADVERSARIAL NETS. 2018.

65. Dumpala, V.; Kurupathi, S.R.; Bukhari, S.S.; Dengel, A.R. Removal of Historical Document Degradations using Conditional GANs. International Conference on Pattern Recognition Applications and Methods, 2019.

66. 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. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* **2017**, pp. 8798–8807.

67. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. *2017 IEEE International Conference on Computer Vision (ICCV)* **2017**, pp. 2242–2251.

68. Kim, T.; Cha, M.; Kim, H.; Lee, J.K.; Kim, J. Learning to Discover Cross-Domain Relations with Generative Adversarial Networks. International Conference on Machine Learning, 2017.

69. 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**, *abs/2010.10103*.

70. Bhunia, A.K.; Bhunia, A.K.; Sain, A.; Roy, P.P. Improving Document Binarization Via Adversarial Noise-Texture Augmentation. *2019 IEEE International Conference on Image Processing (ICIP)* **2018**, pp. 2721–2725.

71. Kumar, A.; Ghose, S.; Chowdhury, P.N.; Roy, P.P.; Pal, U. UDBNET: Unsupervised Document Binarization Network via Adversarial Game. *2020 25th International Conference on Pattern Recognition (ICPR)* **2020**, pp. 7817–7824.

72. Konwer, A.; Bhunia, A.K.; Bhowmick, A.; Bhunia, A.K.; Banerjee, P.; Roy, P.P.; Pal, U. Staff line Removal using Generative Adversarial Networks. *2018 24th International Conference on Pattern Recognition (ICPR)* **2018**, pp. 1103–1108.

73. 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*.

74. Mirza, M.; Osindero, S. Conditional Generative Adversarial Nets. *ArXiv* **2014**, *abs/1411.1784*.

75. Souibgui, M.A.; Kessentini, Y. DE-GAN: A Conditional Generative Adversarial Network for Document Enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2020**, *44*, 1180–1191.

76. De, R.; Chakraborty, A.; Sarkar, R. Document Image Binarization Using Dual Discriminator Generative Adversarial Networks. *IEEE Signal Processing Letters* **2020**, *27*, 1090–1094.

77. Rajesh, B.; Agrawal, M.; Bhuva, M.; Kishore, K.; Javed, M. Document Image Binarization in JPEG Compressed Domain using Dual Discriminator Generative Adversarial Networks. *ArXiv* **2022**, *abs/2209.05921*.

78. 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**, *abs/2211.16098*.

79. Fathallah, A.; El-Yacoubi, M.A.; Amara, N.E.B. EHDI: Enhancement of Historical Document Images via Generative Adversarial Network. VISIGRAPP, 2023.

80. 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.

81. Peng, X.; Wang, C.; Cao, H. Document Binarization via Multi-resolutional Attention Model with DRD Loss. *2019 International Conference on Document Analysis and Recognition (ICDAR)* **2019**, pp. 45–50.

82. Bezmaternykh, P.V.; Ilin, D.; Nikolaev, D.P. U-Net-bin: hacking the document image binarization contest. *Computer Optics* **2019**.

83. 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 Computing and Applications* **2021**, *35*, 13791 – 13802.

84. Ma, K.; Shu, Z.; Bai, X.; Wang, J.; Samaras, D. DocUNet: Document Image Unwarping via a Stacked U-Net. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* **2018**, pp. 4700–4709.

85. Peng, X.; Cao, H.; Natarajan, P. Using Convolutional Encoder-Decoder for Document Image Binarization. *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)* **2017**, *01*, 708–713.

86. 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. *2022 26th International Conference on Pattern Recognition (ICPR)* **2022**, pp. 1699–1705.

87. Chaurasia, A.; Culurciello, E. LinkNet: Exploiting encoder representations for efficient semantic segmentation. *2017 IEEE Visual Communications and Image Processing (VCIP)* **2017**, pp. 1–4.

88. 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.

89. Zhou, L.; Zhang, C.; Wu, M. D-LinkNet: LinkNet with Pretrained Encoder and Dilated Convolution for High Resolution Satellite Imagery Road Extraction. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* **2018**, pp. 192–1924.

90. Ntirogiannis, K.; Gatos, B.; Pratikakis, I. An Objective Evaluation Methodology for Document Image Binarization Techniques. *2008 The Eighth IAPR International Workshop on Document Analysis Systems* **2008**, pp. 217–224.

91. Pratikakis, I.; Gatos, B.; Ntirogiannis, K. H-DIBCO 2010 - Handwritten Document Image Binarization Competition. *2010 12th International Conference on Frontiers in Handwriting Recognition* **2010**, pp. 727–732.

92. Ntirogiannis, K.; Gatos, B.; Pratikakis, I. ICFHR2014 Competition on Handwritten Document Image Binarization (H-DIBCO 2014). *2014 14th International Conference on Frontiers in Handwriting Recognition* **2014**, pp. 809–813.

93. Pratikakis, I.; Zagoris, K.; Barlas, G.; Gatos, B. ICFHR2016 Handwritten Document Image Binarization Contest (H-DIBCO 2016). *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)* **2016**, pp. 619–623.