Document image binarisation: an overview and a practical comparison study

Mawloud Mosbah

Document image binarisation: an overview and a practical comparison study

Mawloud Mosbah

Informatics Department,
Faculty of Sciences,
University 20 Août 1955 of Skikda,
Skikda, Algeria
and
LRES Laboratory,
University 20 Août 1955 of Skikda,
Skikda, Algeria
Email: ma.mosbah@univ-skikda.dz
Email: mos_nasa@hotmail.fr

Abstract: In this paper, we deal with document image binarisation, as an essential image pre-processing step, through presenting a comparison study of the different methods of the literature in terms of effectiveness. In addition, this sub-field of document image binarisation, including a lot of methods, techniques, and algorithms, proposed along the recent decades of years, is overviewed here through categorising the various methods into their different approaches as well as talking about the evaluation step via showing the metrics and the considered public datasets and benchmarks. Some issues caused by different types of noise, some frequent utilised filters, considered as pre-processing, as well as some other techniques of post-processing are also presented. To note that the results compared here are as given in the different references that are either regular papers or comparative studies in the form of competitions.

Keywords: document image; document image binarisation; image binarisation.

Biographical notes: Mawloud Mosbah is an Associate Professor at Informatics Department, University 20 Août 1955 of Skikda and member of LRES Research Laboratory within the same university. He has received his Doctor of Sciences and Habilitation in informatics from University 20 Août 1955 of Skikda respectively in 2017 and 2022. His main areas of interest and expertise include information retrieval, image retrieval and automatic natural language processing.

1 Introduction

Due to the widespread and the relative low price of new technological tools including scanners and cameras that capture images such as smart phones, it is so easy, for everyone, to photograph any scene he/she meets. Although other wired and wireless

alternatives for information exchange and communication, capturing the scene with camera seems to be very simple and easy.

Capturing text documents, with cameras, generates what we know as a (print-written) document image. These document images may be also generated when we scan historical archive documents as handwritten document images. There are many challenges addressed especially in handwritten document image binarisation, such as faint characters, bleed-through, smudges, non-uniform illumination, and large background ink stains (Mustafa and Kader, 2018a). Each of tools for generating document image, either a camera or a scanner, has its proper problems. Unfortunately, we cannot treat a document image, may be open in an image editor, as the original text document, may be open in a text editor, where we can search, copy, and paste words, phrases, and paragraphs. All the text, in a document image, is considered unfortunately as just a set of pixels.

Converting a document image to its original text aims to treat, process, and manipulate the text automatically, by the machine, taking profit of operations furnished by any simple text editor and any tool for automatic processing of natural language. The global context of this problem is then a document image recognition and understanding.

Consulting of literature reveals the following:

- A lot of works that address document image binarisation especially in the last few decades.

- Unfortunately the secondary publications in terms of overviews, reviews, and states of arts are rare. These rare publications focus on theoretical aspect without any comparison of what exists as techniques, methods, and tools.

- There are some competitions that compare between some works in the field without including the necessary theoretical and academic aspect.

This work addresses then the sub-field of document image binarisation, trying to mix between the theoretical and the practical aspects, through:

1 categorising the different methods, techniques and tools

2 highlighting the problems of the images before binarisation

3 shedding the light on the various filters as pre-processing schemes.

From the practical point of view, a comparison of the methods is made based on what exists in the literature, as a capitalisation, without conducting any personal experimentation.

The rest of the paper is organised as follows: Section 2 presents the approaches and their basic methods of the literature. Section 3 deals with noises may degrade image documents and filters may be used in order to enhance the quality of the image. Section 4 talks about the evaluation process while Section 5 shows the experimental results as given in the different considered references.

2 Document image binarisation: approaches and methods

One of the most crucial steps of pre-processing of document images subjected to further text recognition is their binarisation, which influences significantly obtained OCR results

(Krupinski et al., 2019). By definition, the binarisation (also known thresholding or foreground-background separation) of an image means labelling its text as text (foreground) or background. Background is, by its self, defined very vaguely, and it may contain unwanted information which may even carry some form of textual data. The motivation of research related with document image binarisation and recognition is not only the possibility of preserving the cultural heritage and discovering some historical facts but also potential applications of the developed algorithms in some other areas of industry (Krupinski et al., 2019). Document image binarisation is then an active area of research in computer vision where many algorithms have been proposed throughout the literature. These algorithms can be divided into two major categories: clustering-based methods (Moghaddam and Cheriet, 2010) and threshold-based methods (Moghaddam and Cheriet, 2010) and may be hybrid methods such as Kwon (2004). Clustering-based methods usually use model-based features to differentiate text from background and classify it (Moghaddam and Cheriet, 2010). Fuzzy classification (Papamarkos, 2001) and recursive segmentation in the PCA space (Drira, 2006) are two examples of clustering-based methods. In the threshold-based category, the most famous method, and a pioneering one, is Otsu's method (Chaki et al., 2014), which is based on maximising the separation between two pre-assumed classes. A previous work (Abd Elfattah et al., 2015) classified the methods into six categories: histogram-based methods, clustering-based methods, entropy-based methods, foreground attribute-based methods, spatial binarisation methods, and locally adaptive methods. It is worthy to note that there are other methods coming from novel disciplines like optimisation such as Abd Elfattah et al. (2015) in which artificial bee colony (ABC) algorithm has been invoked as a clustering algorithm in order to segment the text from the images. The ABC algorithm helps the proposed algorithm to find the best threshold value to get high quality binarised image according to performance evaluation measures.

Consulting literature reveals that there are some works of second position such as: surveys, reviews, and overviews that deal with image binarisation (Mustafa and Kader, 2018a; Chaki et al., 2014; Bhowmik, 2023; Mustafa et al., 2018; Chauhan et al., 2016; Sulaiman et al., 2019; Tensmeyer and Martinez, 2020). In Mustafa and Kader (2018a), the authors have provided a comprehensive review, discussed, and tested seven types of binarisation method on *H-DIBCO12*. As a conclusion, they have given some critics towards the considered techniques of the literature, in the hope to improve performances, like considering traditional methods for contrast enhancement as a post-processing step. In Bhowmik (2023), the author has considered three major groups for document image binarisation, namely: threshold-based approach, optimisation-based approach, and classification-based approach. Although that Chaki et al. (2014) did not present any information about evaluation, they have given an essential idea about the assessment step which is indirect evaluation saying that the quality of any binarisation method may appear later in the degree of effectiveness of the recognition applications coming later that utilise the output of the binarisation as an input. Away from the classical categorisation of binarisation methods considered in the majority of works, Chaki et al. (2014) divide binarisation methods, based on their chronological appearance, into two categories: foundations and recent works. In Mustafa et al. (2018), the authors have presented a comparison of several image binarisation techniques according to *H-DIBCO 2013* dataset. In Chauhan et al. (2016), the authors have presented a review and a comparison study of some various document image binarisation techniques

according to *DIBCO-2009* and *DIBCO-2010* datasets. In Sulaiman et al. (2019), the authors have considered four categories for document image binarisation, namely:

1 global thresholding-based binarisation methods

2 local/adaptive thresholding-based methods

3 hybrid thresholding-based methods

4 machine learning-based methods.

In Tensmeyer and Martinez (2020), the authors have provided a comprehensive review of the field of historical image binarisation. They have presented a table which summaries and categorises the various techniques introduced in the last decade. Among these categorisations, we can quote:

1 global threshold-based techniques

2 local threshold-based techniques

3 edge-based techniques

4 image transform-based techniques

5 mixture model-based techniques

6 conditional random fields-based techniques

7 game theory-based techniques

8 shallow machine learning-based techniques

9 deep learning-based techniques

10 supervised tuning-based techniques

11 unsupervised tuning-based techniques.

In this work, we consider four approaches for document image binarisation:

1 image binarisation based on clustering

2 image binarisation based on threshold

3 binarisation based on neural networks

4 combination of binarisation techniques.

2.1 *Image binarisation based on clustering*

Clustering and classification are both fundamental tasks in data mining. Classification is used mostly as a supervised learning method, clustering for unsupervised learning. There are some works of literature where clustering (unsupervised classification) algorithms are used for document image binarisation.

2.1.1   *Binarisation based on k-means*

Some works have considered k-means for document image binarisation (Jana et al., 2017a). Indeed, in Jana et al. (2017a), the authors have proposed a k-means-based clustering technique for adaptive binarisation of degraded document images. Figure 1 presents an original image and its binarised version using K-means.

**Figure 1**   Original image and its binarised version using k-means (see online version for colours)
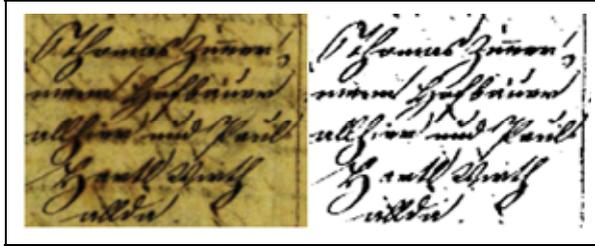
Source: Jana et al. (2017)

2.1.2   *Binarisation based on fuzzy classification*

Many works have used fuzzy C-means method (FCM) in the binarisation of document images such as in Papamarkos (2001), Mustafa et al. (2019a), Jana et al. (2017b), Mondal et al. (2019) and Tong et al. (2009). Fuzzy classification is generally considered with local thresholding methods in order to fasten them. Indeed, in Papamarkos (2001), the authors have proposed a method for fuzzy binarisation of digital document claimed suitable for binarisation of blurring or badly illuminated documents and can be easily modified to accommodate any type of spatial characteristics. After training, the neurons of the output competition layer of the considered self-organised feature map (SOFM) define two bi-level classes. Using content of these classes, fuzzy membership functions are obtained that are next used with the FCM algorithm in order to reduce the character-blurring problem. In Mustafa et al. (2019a), the proposed consists of three parts. In the first part, the image pre-processing operation has been done before the binarisation process to enhance image quality through applying both contrast stretching and mean filter. The second part is to apply the binarisation algorithm on the document image that has undergone an image pre-processing operation. By applying the FCM algorithm to the document images, the images are converted to binary images and divided into two components, which is text and background. The last step of the proposed method is to perform deghost operation to remove 'ghost' entities that may have appeared on the document image. In Jana et al. (2017b), the authors have presented a fast and competent, yet simple binarisation technique that uses a FCM-based global thresholding approach, aided by background separation. The proposed method uses a superset of foreground regions to correctly assess background of the document image. Background is estimated based on a sliding interpolation window of variable dimension, judged by appraising the nature of text stroke. Ultimately, a global approach is undertaken to binarise the background-separated normalised and enhanced image by clustering the pixels using FCM. In Mondal et al. (2019), the authors have proposed a local threshold binarisation method using fast fuzzy C-means clustering. Historical document images with non-uniform background, stains, faded ink are first processed by removing the

background using in-painting-based method. Then FCM clustering is used to cluster out the pixels into three main clusters: sure text pixels, sure background pixels and confused pixels which may or may not be labelled as text. In Tong et al. (2009), the authors have proposed a binarisation algorithm, called NFCM, which is a combination of Niblack algorithm and FCM algorithm. They have claimed that their algorithm is good not only at preserving the character stokes, but also at alleviating the ghost artifacts. Figure 2 gives some images from HDIBCO 2013 binarised using FCM method.

Figure 2   Some images from HDIBCO'13 binarised using fuzzy C-means method
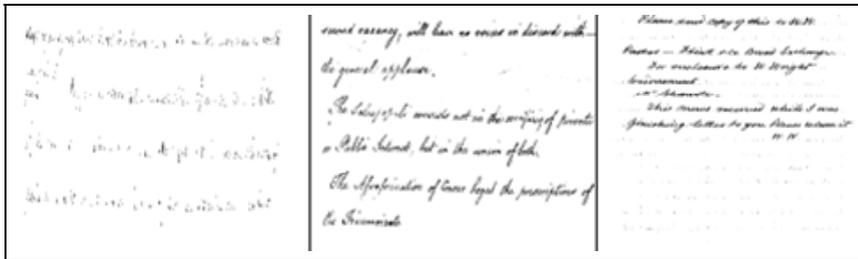
2.2   *Image binarisation based on threshold*

Image binarisation consists commonly to convert a colour or grey image into a binary image, where the text and background pixels are marked in black and white respectively (Jia et al., 2018). A successful binary result preserves meaningful information while discarding noisy information (Wu et al., 2016).

Thresholding is a simple yet powerful technique to separate the object of interest from the background (Yazid and Arof, 2013). Image thresholding has found various applications in many computer vision and graphics applications such as document image analysis. In an image, if the objects are clearly lighter (or darker) than the background it is natural to separate them by setting a threshold (Yazid and Arof, 2013).

The algorithm of binarisation-based threshold is simple. Given a fixed global threshold value T (from 0 to 255), if the intensity value of any pixel of an input image is more than T, the pixel is set to white otherwise it is black. If the source is a colour image, it first has to be converted to grey level.

Unfortunately, a fixed global value for the threshold may degrade drastically the quality of the image, owing to the variation of luminosity over the different regions of the input image generating then many misclassified pixels. This fact leads researchers to think about considering either a dynamic threshold (Bernsen, 1986), an adaptive threshold as done in Tensmeyer and Martinez (2020), or even considering a multiple local thresholds for the same image (Jia et al., 2018; Kim, 2004). Due to the importance of thresholding (Sezgin and Sankur, 2004), binarisation methods are categorised depending on which principal criteria they consider in calculating the threshold (Chaki et al., 2014). The core problem, tied to binarisation-based threshold, to be asked then, is how to designate the threshold.

Consulting the literature reveals a lot of works to select the binarisation threshold for any image in general and image document implementing optical character recognition (OCR) tools in particular. There is the method based on clustering-analysis (Otsu, 1979),

methods based on entropy (Johannsen, 1982; Kapur et al., 1985), methods based on image variance (Sauvola and Pietikainen, 2000; Niblack, 1985), method based on image contrast (Wu et al., 2016), methods relying on stroke structural symmetry of strokes (Drira, 2006), and those based on texture features (Liu and Srihari, 1997). In He et al. (2005), the authors have conducted a comparison for some binarisation methods for historical archive documents.

It is worthy to note that for images with poor and non-uniform illumination, adaptive thresholding is required to separate the objects of interest from the background (Yazid and Arof, 2013; Bloechle et al., 2024).

2.2.1   Global thresholding

In global thresholding approach, a single threshold value is determined to be applied to the whole image in order to put aside the pixels under consideration into foreground and background. In the following sub-sections, we quote the basic global-threshold binarisation methods.

### 2.2.1.1   Gradient-based thresholding

In Yazid and Arof (2013), a new approach to create an adaptive threshold surface is proposed to segment an image. The technique is inspired by the Yanowitz's method (Yanowitz and Bruckstein, 1989) and is improved upon by the introduction of a simpler and more accurate threshold surface. The method uses gradient-based thresholding by constructing a threshold surface. It contains three important phases: first, construct the inverse image $T(i, j)$, second obtaining the k value between [–255, 255], and finally applying binarisation to separate object and background.

$$result(i, j) = \begin{cases} 0, & (object) & if\ I(i, j) < T(i, j) + k_0 \\ 255, & (background) & if\ I(i, j) > T(i, j) + k_0 \end{cases} \tag{1}$$

where $I(i, j)$ is the intensity of the original image and k_0 is the minimum sum of absolute difference intensity.

Figure 3 depicts some images from HDIBCO 2016 binarised using gradient-based threshoding method.

Figure 3   Some images from HDIBCO'16 binarised using gradient-based thresholding method
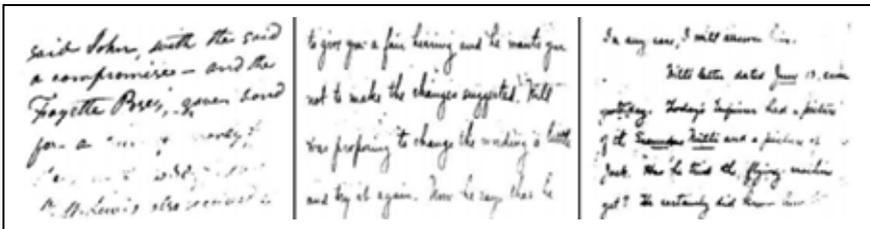
2.2.1.2 Otsu method

In Otsu (1979), a very simple method with an optimal threshold and straightforward extension to multi-thresholding problem is introduced. The threshold is obtained automatically based on global variance and between-class variance. In the non-uniform image, Otsu assumes the image contains two areas: dark and bright in order to purpose final algorithm. Otsu thresholding is determined then by:

$$k = \frac{\sigma^2 B}{\sigma^2 G} \tag{2}$$

where K is a threshold value, $\sigma^2 B$ is a global variance of the entire image, and $\sigma^2 G$ is a between-class variance.

Figure 4 presents some images from HDIBCO 2016 binarised using Otsu method.

Figure 4   Some images from HDIBCO'16 binarised using Otsu method
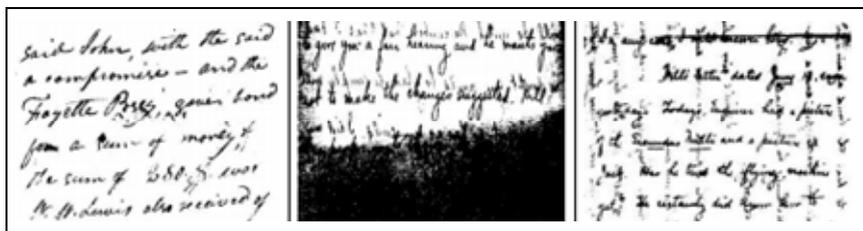
2.2.2 Local thresholding

Unfortunately, global threholding is not sufficient in some situations (Sauvola and Pietikainen, 2000) caused by changes in lamination (illumination), scanning errors and resolution, poor quality of the source document and complexity in the document structure. In case of local or adaptive thresholding, same threshold is never used throughout the entirety of the image. Rather, the properties of a pixel and its neighbours in a sub-image help to determine the threshold. In the following sub-sections, we give some basic local-threshold binarisation methods.

2.2.2.1 Niblack method

The main purpose of Niblack method (Niblack, 1985) is to set the threshold value based on local standard deviation and local mean. The threshold for each pixel is determined by:

$$T(x, y) = m(x, y) + k\delta(x, y) \tag{3}$$

where local standard deviation $\delta(x, y)$ and local mean $m(x, y)$ are determined by 80×80 windowing size and standard k value is (–0.2). It is worthy to note that this method does not work correctly if the image suffers from non-uniform illumination. Figure 5 gives some images from HDIBCO 2016 binarised using Niblack method.

Figure 5   Some images from HDIBCO'16 binarised using Niblack method
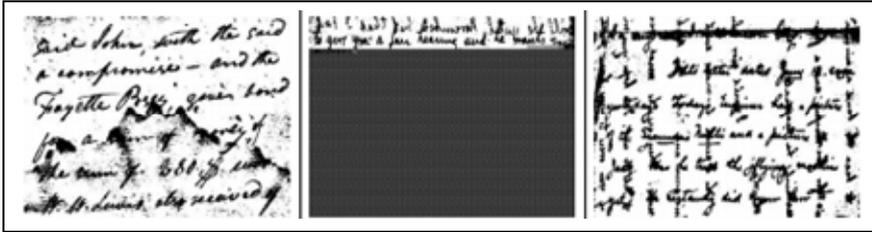
2.2.2.2   *Sauvola method*

In Sauvola and Pietikainen (2000), the authors have introduced a new method for document image binarisation using an adaptive approach to manage different situations in an image. Sauvola technique uses rapid image surface analysis for algorithm selection and adaptation according to document contents. The content is used to select the algorithm type ad need for parameterisation and to compute and propose the threshold value for each or every nth pixel (interpolative approach). The document content is used to guide the binarisation process. Figure 6 depicts an image from HDIBCO 2012 and its associated binarised image using Sauvola method.

Figure 6   Image from HDIBCO'12 and its associated binarised image using Sauvola method
　　　　　　(see online version for colours)
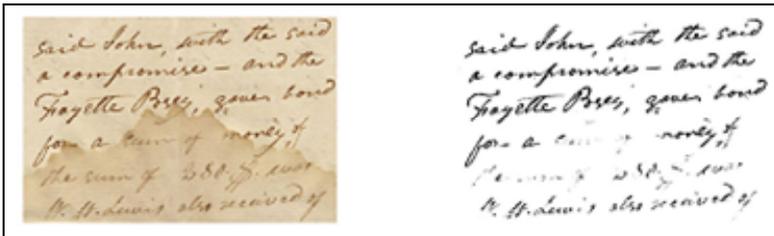
Figure 7   Some images from HDIBCO'13 binarised using deghost method
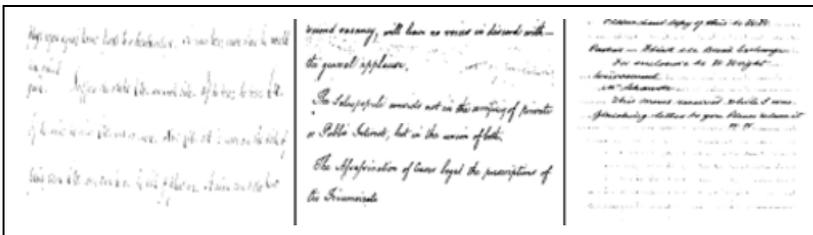
2.2.2.3   *Deghost method*

In Li et al. (2018), the authors have proposed a learning rate rule that combines adaptive threshold edge detection and a temporal gate. Through the noise estimation algorithm, the adaptive spatial threshold is related to the residual non-uniformity noise in the corrected image. The proposed learning rate is used to effectively and stably suppress ghosting

artifacts without slowing down the convergence speed. Figure 7 presents some images from HDIBCO 2013 binarised using deghost method.

2.2.2.4 Nick method

In Khurshid et al. (2009), a sliding window-based local thresholding technique, trying to improve the Niblack method by shifting down the binarisation threshold, is presented. The threshold is obtained based on the following equation:

$$T(x, y) = m + k\sqrt{\frac{(I^2 - m^2)}{N}}$$

(4)

where the *k* factor is similar with Niblack and the windowing size is defined as 15 × 15, while *I* and *m* represent the intensity pixel and mean of greyscale image. *N* represents the image size. Figure 8 gives some images from HDIBCO 2016 binarised using Niblack method.

Figure 8 Some images from HDIBCO'16 binarised using Niblack method
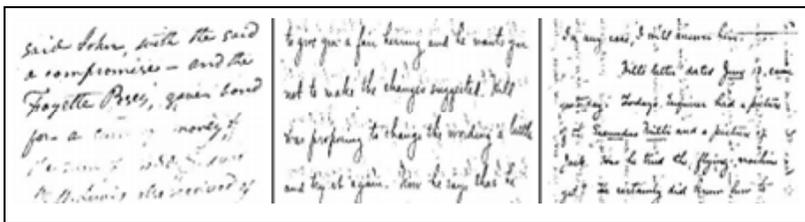
Figure 9 Some images from HDIBCO'16 binarised using Bradley method
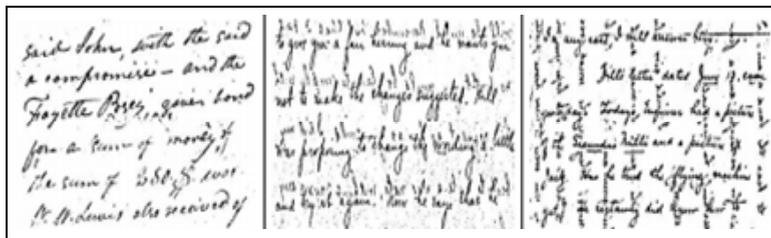
2.2.2.5 Bradley method

In Bradley and Roth (2007), the authors have presented an improvement and an extension of Wellner's method (Pierre, 1993) which is robust to illumination changes within the image. The main idea in Wellner's algorithm is that each pixel is compared to an average of the surrounding pixels while the key idea of the Bradley algorithm is that every image's pixel is set to black if its brightness is *T* percent lower than the average brightness of surrounding pixels in the window of the specified size, otherwise it is set to white. The default windowing size is 15 × 15 and *T* is 10. T is given as follows:

$$T = m\left(1 - \frac{k}{100}\right)$$

(5)

Figure 9 presents some images from HDIBCO 2016 binarised using Bradley method.

2.2.2.6 *Bernsen method*

Bernsen algorithm (Bernsen, 1986) is based on the estimation of a local threshold value for each pixel. This value is assigned the local threshold value only if the difference between the lowest and the highest grey level value is bigger than threshold k. Otherwise; it is assumed that the window region contains pixels of one class (foreground or background). The default windowing size (w) is 3×3 and k is 15. The final equation as follows:

$$T(x, y) = \frac{Z_{max} + Z_{min}}{2} \tag{6}$$

where Z_{min} and Z_{max} are the lowest and highest grey level pixel values. Figure 10 depicts some images from HDIBCO 2016 binarised using Bernsen method.

Figure 10   Some images from HDIBCO'16 binarised using Bernsen method
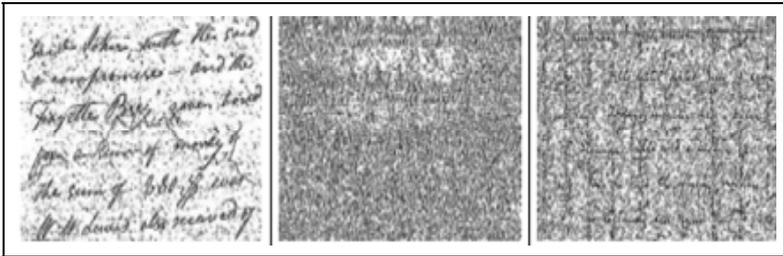
Figure 11   Some images from HDIBCO'16 binarised using local adaptive method
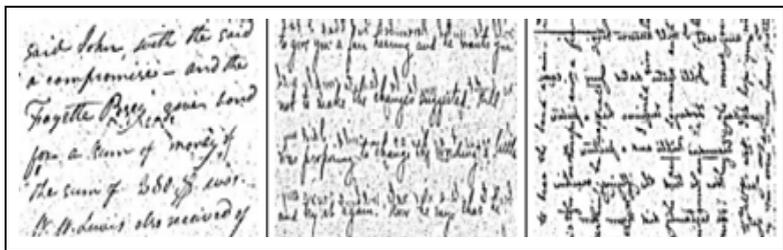
2.2.2.7 *Local adaptive thresholding*

Local adaptive thresholding (Woods and Gonzalez, 2008) is a basic simple algorithm to separate the foreground from the background with non-uniform illumination. For each pixel in the image, a threshold has to be calculated. If the pixel value is below the threshold, it is set to the background value, otherwise, it assumes the foreground value. The default local windowing size (w) is 15×15 and local threshold (T) is a 0.05. The threshold is given as follows:

$$T = \frac{max + min}{2} \tag{7}$$

Figure 11 depicts some images from HDIBCO 2016 binarised using local adaptive method.

2.2.2.8 Gatos method

In Gatos et al. (2006), the authors have proposed an adaptive approach for the binarisation and enhancement of degraded documents. They claimed that the proposed method does not require any parameter tuning by the user and can deal with degradations. The approach is an entire process which includes three essential tasks: pre-processing, processing, and post-processing. The pre-processing, which aims to eliminate noisy areas, smoothing of background texture as well as contrast enhancement between background and text areas, considers an adaptive Wiener filter. After that an initial segmentation of foreground and background regions as well as the background surface estimation is done. A final thresholding is performed then by combining the calculated background surface with the pre-processed image. A post-processing is considered as a final step in order to eliminate noise, improve the quality of text regions, and preserve stroke connectivity by isolated pixel removal and filling of possible breaks, gaps or holes. Figure 12 presents an image from DIBCO 2009 with its binarised image using Gatos method.

Figure 12 Image from DIBCO'09 and its binarised image using Gatos method
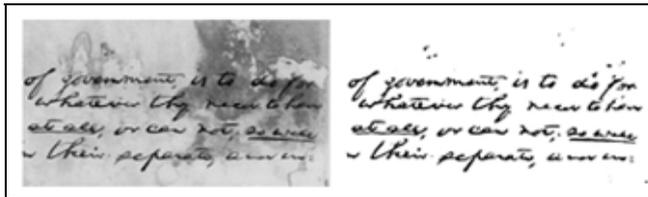
2.3 Binarisation based on neural networks

Some works such as Tensmeyer and Martinez (2017), Kefali et al. (2014), Sari et al. (2012) and Akbari et al. (2020) have adopted neural networks as an artificial intelligence tool for dealing with document image binarisation. Indeed, in Tensmeyer and Martinez (2017), the authors have formulated binarisation issue as a pixel classification learning task, to address a multiple image scales, including full resolution, where a fully convolutional network (FCN) architecture has been applied. In Kefali et al. (2014), a foreground-background separation technique, based on artificial neural networks (ANN), has been applied on old documents with a variety of degradation. The idea has been to train an ANN on a set of pairs of original images and their respective ideal black and white ones relying on global and local information. In Sari et al. (2012), the authors have considered a back-propagation neural network to directly classify image pixels according to their neighbourhood. In Akbari et al. (2020), the authors have proposed the utilisation of convolutional neural networks (CNN) to identify foreground pixels using novel input-generated multichannel images. To create the images, the original source image is decomposed into wavelet sub-bands. Then, the original image is approximated by each sub-band separately, and finally, the multichannel image is constituted by arranging the original source image as the first channel and the approximated image by each sub-band as the remaining channels. Two scenarios are considered, that is, two-channel and

four-channel images, and then fed into two types of CNN architectures, namely, single and multiple streams. CNNs used in the architectures are the three popular networks: U-net, SegNet, and DeepLabv3+. Figure 13 depicts an original image and its binarised version using fully convolutional neural networks (FCNN).

Figure 13   Original image and its binarised version using fully convolutional neural networks (FCNN)
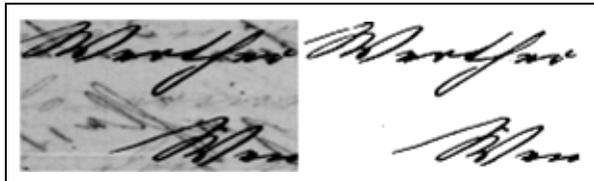
Source: Tensmeyer and Martinez (2017)

2.4   *Combination of binarisation techniques*

Some works of literature have preferred to consider combination of some existed methods (Badekas and Papamarkos, 2007; Sari et al., 2014; Yoon et al., 2013; Lelore and Bouchara, 2013) of binarisation rather than of introducing new binarisation technique. Indeed, in Badekas and Papamarkos (2007), the authors have built a system that takes advantages of the benefits of a set of selected binarisation techniques by combining their results using a Kohonen self-organising neural network. They claimed that the built system is suitable to classify pixels that have high vagueness. That is, pixels which belong to edges, shadow areas and generally pixels that cannot be easily classified as foreground or background pixels. In Sari et al. (2014), the authors have presented a technique based on hybrid thresholding combining the advantages of global and local methods. Two stages have been included. In the first stage, global thresholding is applied on the entire image and two different thresholds are determined from which the most of image pixels are classified into foreground or background. In the second stage, the remaining pixels are assigned to foreground or background classes based on local analysis. In this stage, several local thresholding methods are combined and the final binary value of each remaining pixel is chosen as the most probable one. In Lelore and Bouchara (2013), the authors have introduced the FAIR algorithm for document image restoration. They claimed that the algorithm is scale invariant and fast enough to be used in real-time applications. It is based on a double-threshold edge detection approach as well as clustering which makes it possible to detect small details while remaining robust against noise. Figure 14 gives the binarisation of some images from HDICO 2013 using multiple thresholds.

Figure 14   Some images from HDIBCO'13 binarised using multiple thresholds
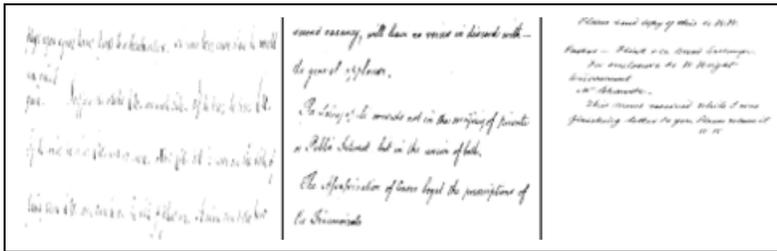
Source: Mustafa et al. (2019a)

It is worthy to note that there are some approaches that go beyond the scope of that given classification such as the method introduced by Howe (2013) which mainly relies on minimising the global energy function inspired by a Markov random field. Figure 15 depicts an original image and its binarised version using Howe method.

Figure 15   Original image and its binarised version using Howe method
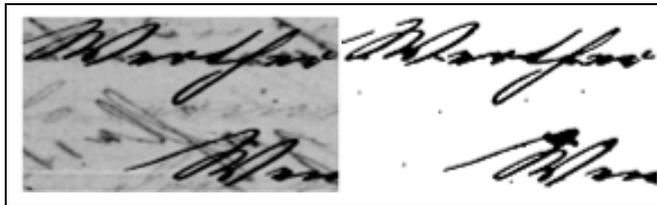
3   Noises and filters

3.1   Document image degradation and noises

Bhowmik (2023) has categorised document image degradation and noise into two categories:

1   noise from native sources which appears due to aging, quality of paper and ink used

2   noise coming from auxiliary sources or external factors like malfunctioning of scanning devices, lighting condition during acquisition, misplacement of the document during scanning.

The noise coming from native sources causes issues like:

- *Ink-bleeding or bleed through:* which appears when both sides of the page are used to write and the ink seeps through one side and spreads over to the other side.

- *Show through or smear:* which appears because of the ink impression of one side appears on the other side. It often creates dark spots on the page.

- *Faint text:* it appears in documents written or printed using low quality ink. This is because as the time grows the ink starts to shrink. The quality of paper used may also be the reason for that.

- *Deterioration of document:* takes place due to aging, poor storage, mishandling, natural calamity and other environmental condition. This type of noise causes presence of dark spot and other artefacts on the document images.

The noise coming from auxiliary sources causes problems such as:

- *Uneven illumination:* which occurs in the light microscopy images as in optical imaging where the incident light decreases drastically along the path.

- *Contrast variation:* this appears due to the environment under which the image acquisition process is carried out.

- *Blurring effect:* observed in document images are of two types: motion blurring, appearing due to the relative motion of the camera and object or sudden rapid movement of the camera, and out-of-focus blurring, appearing where some of the points are in focus while others are not during acquisition.

- *Skew:* this appears either due to the misplacement of the document during scanning or due the writing style. This can occur either at the page level or at the region level or both. Generally, for the printed documents, page level skew is found whereas in handwritten documents region level skew is observed mostly.

More information about degradation problems and their main models are given in Zhou et al. (2023).

3.2 *Some filters as pre-processing, processing and post-processing steps for binarisation*

Pre-processing consists of preparing the document image that suffers from many problems such as shadow, noisy areas from scanning, document aging, ink blot, non-uniform illumination, ink seepage, smudges, smear, strain, bleed-through, background ink-stain, and fading while post-processing, considered commonly as a final step of the binarisation process, consists of eliminating eventual remaining noise such as ghosts, improving the quality of text regions, constructing the broken lines, and filling of possible breaks, gaps or holes.

There are a lot of filters considered as pre-processing or post-processing steps for document image binarisation operation. In this section, we quote the essential ones that are frequently used.

3.2.1 *Gabor filter*

In Sehad et al. (2019), the authors have presented a simple method suitable for a software and hardware implementation based on Gabor filters for the binarisation of ancient degraded documents. The Gabor filter bank is designed by taking into account the degradation type the document based on the un-blind protocol. First, the document image is pre-processed using a Wiener filter to smooth the degradation. Subsequently, the binarisation threshold is estimated using texture features, such as the mean and the standard deviation, extracted from the respective original image and the filtered document image. Furthermore, a new protocol, namely un-blind protocol, is proposed for estimating the standard deviation according to the degradation type for setting the optimal parameters of the Gabor filter such as the central frequency and the number of angles.

Figure 16 depicts an original image and some associated Gabor Filtered images with different parameters.

Figure 16   (a) Original image (b) Gabor filtered image with $\sigma < \rho$ (c) Gabor filtered image with $\sigma = \rho$ (d) Gabor filtered image with $\sigma > \rho$

(a)

(b)

(c)

(d)

3.2.2   Generalised Gaussian distribution

In Krupinski et al. (2019), the authors have confirmed the use of generalised Gaussian distribution (GGD) histogram approximation with the use of the Monte Carlo method for pre-processing of degraded document images before binarisation and further analysis. Indeed, through extracting the parameters of GGD, distortions may be modelled and removed, enhancing the quality of input data for further thresholding and text recognition. Due to an efficient use of the Monte Carlo method, the overall processing time has been shortened and the proposed approach-based GGD may be further combined with other binarisation algorithms not considered in Krupinski et al. (2019). Figure 17 gives a scanned document image and its version with Gaussian noise.

Figure 17   Scanned document image and its version with Gaussian noise

3.2.3   Wiener filter

In Gatos et al. (2006), the authors have used adaptive *Wiener filter*, based on local statistics estimated from a local neighbourhood around each pixel, as a pre-processing step to de-noise the document image. The use of a low-pass *Wiener filter* (Jain, 1989) has proved efficient for the elimination of noisy areas, smoothing of back-ground texture as well as contrast enhancement between background and text areas. The *Wiener filter* is commonly used in filtering theory for image restoration. Figure 18 depicts an image and its associated Wiener filter.

Figure 18   An image and its associated Wiener filter
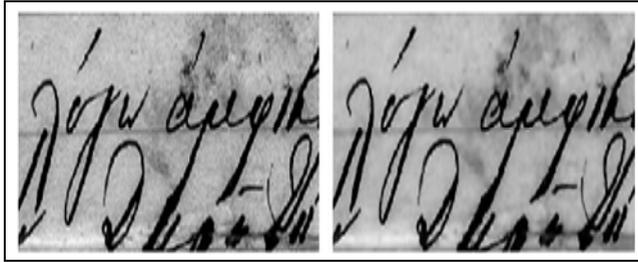
3.2.4 Shrink and swell filtering

In Gatos et al. (2006) and Biswas et al. (2014), the authors have considered shrink and swell filtering as a post-processing. A shrink filter is used to remove noise from the background through converting a foreground pixel surrounded by background pixels to a background pixel while a swell filter is used to fill possible breaks, gaps or holes in the foreground via assuming a background pixel among foreground pixels as a hole in the object to be filled up with the foreground pixel value. Figure 19 gives a binarised image and its version after a post-processing considering swell and filtering.

Figure 19   Binarised image and its version after a post-processing considering swell and filtering
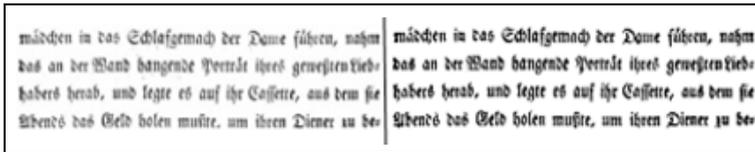
Figure 20   An original image and its version after a local entropy filter

3.2.5 Local entropy filter

In Michalak and Okarma (2019), the authors have considered the local entropy filter as one of the pre-processing steps for adaptive image binarisation of unevenly illuminated document images subjected to further optical text recognition. Figure 20 presents a document image and its associated version after considering local entropy filter.

3.2.6 Median filter

In Nafchi et al. (2013), Gaussian and median filtering are used in order to improve the final binarised output. Gaussian filter is used for further enhancing the output and median filter is applied to remove noises. Figure 21 depicts an image and its version using a method combined with median filter.

Figure 21   An original image and its version using a method combined with median filter
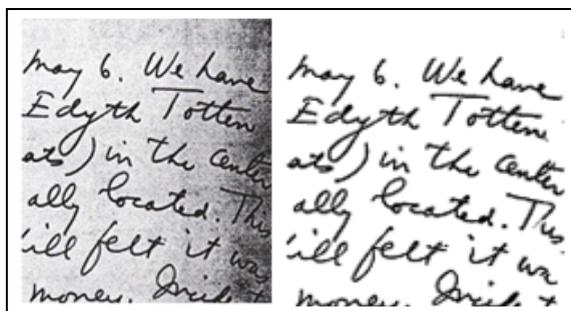
3.2.7 Bilateral filter

In Almeida et al. (2018), the authors have considered bilateral filter which first introduced by Aurich and Weule (1995) under the name 'nonlinear Gaussian filter'. It was later rediscovered by Tomasi and Manduchi (1998) who called it the 'bilateral filter'. The bilateral filter is a technique to smoothen images while preserving their edges. The weight assigned to each neighbour decreases with both the distance values among pixels of the image plane (the spatial domain S) and the distance on the intensity axis (the range domain R). Figure 22 gives a document image and its binarised version using a method with bilateral version.

Figure 22   An original image and its binarised version using a method with bilateral filter
             (see online version for colours)
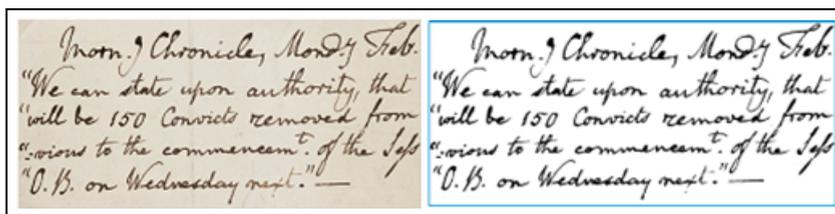
Source: Almeida et al. (2018)

3.2.8   Kalman filter

In Rahiche and Cheriet (2019), the authors have proposed a two-step approach, called KFBin, for the binarisation of document images based on the Kalman filtering technique. According to Chang and Chang (2017), Kalman filter is probably the most widely used technique in statistical signal processing, that processes data samples recursively in a very effective manner. Kalman filter is used in the literature for image de-noising, image reconstruction, video de-noising, and object tracking. In Rahiche and Cheriet (2019), the authors have used Kalman filter to map any multi-channel input images to two outputs images, whereby one output contains the enhanced foreground component and the other contains the background component. Figure 23 presents an image from DIBCO 2014 and its associated version after considering Kalman filter.

Figure 23   An image from DIBOC'14 and its associated version after Kalman filter (see online version for colours)
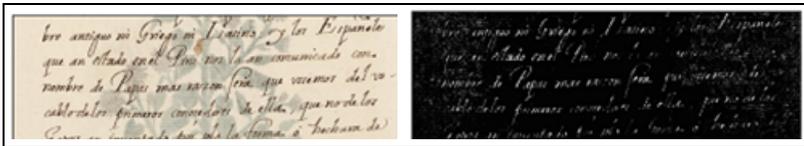
3.2.9   Swell filter

In Chen and Wang (2017), the authors have used swell filter, as a post-processing step, to scan the entire binary image and each white pixel is examined.

Unfortunately and to the best of our knowledge, some interesting filters, such as guided filtering (He et al., 2012) and Bayesian filter (Sarkka and Svensson, 2023), have not been well considered, up to now, for document image binarisation. Guided filtering, as bilateral filter, shares the property of edge-preserving smoothing with a better behaviour near the edge (Bera et al., 2021). To note that in Su et al. (2010) and Pratikakis et al. (2018), where many methods of the literature have been compared, many filters such as median filter, Gaussian filter, guided filter, bilateral filter, and Wiener filter have been considered by the methods made in competition. Figure 24 depicts a binarised image and its version using swell filter as a post-processing task.

Figure 24   Binarised image and its version using swell filter
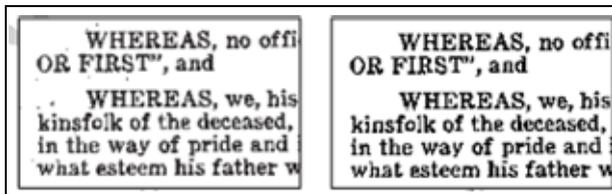
3.2.10   Curvelet transform

In Wen et al. (2013), the curvelet transform and Otsu's method were combined to binarise the non-uniform illuminated images. The non-uniform illumination image was

decomposed by the curvelet and the curvelet coefficients were enhanced by nonlinear functions. The reconstructed image was processed by Otsu's binarisation method, as given in Figure 25.

Figure 25 An image and its binarised version using Otsu method and curvelet transform
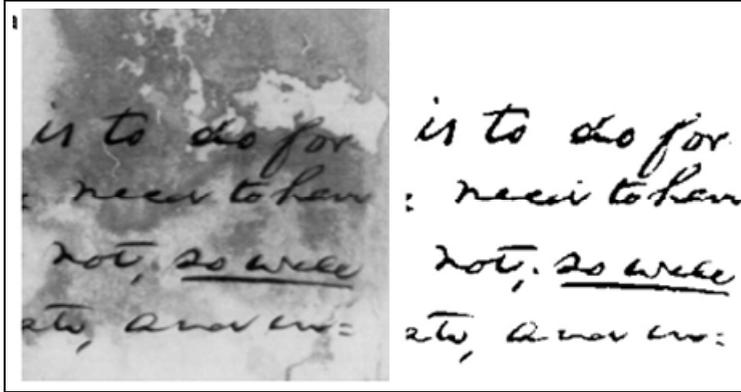
3.2.11 *Contrast stretching filter*

Contrast stretching (often called normalisation), as given in Figure 26, is a simple image enhancement technique that attempts to improve the contrast in an image by stretching the range of intensity values it contains to span a desired range values. In Abd Elfattah et al. (2015), the authors have considered, as a first phase, stretching the intensity level of the image by contrast stretching filter and removing the noise by image cleaning algorithm.

Figure 26 An orginal image and its contrasted version (see online version for colours)
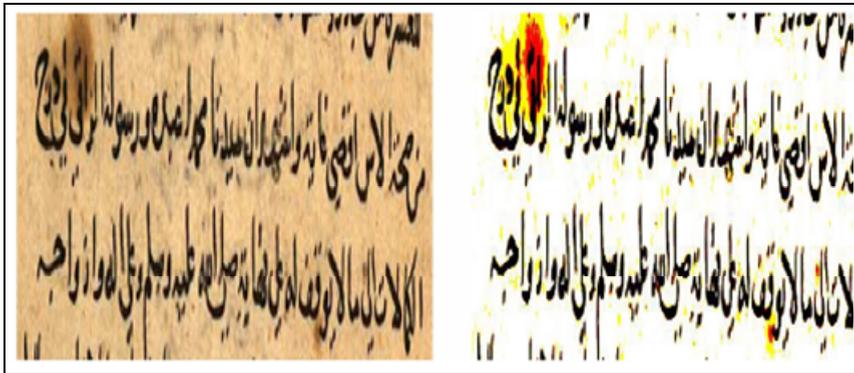
4 Evaluation of binarisation

In this section, we deal with evaluation in binarisation through presenting the different metrics considered in the literature as well as the various considered datasets.

4.1 Metrics

There exist a lot of metrics to evaluate the performance of a binarisation system. Each metric is useful to evaluate one aspect of quality from one side. Therefore, each one should be used alongside the others to get a comprehensive evaluation of the binarisation quality. In this section, we present the essential metrics may be used in the literature.

4.1.1 Precision

Accuracy-based metrics like: precision, recall, F-measure, and accuracy are often the first indicators of how well a binarisation method has performed. The accuracy-based metrics are based on some fundamental metrics of classification of pixels.

Precision value, which measures the proportion of correctly identified foreground pixels among all pixels identified as foreground, is calculated through the true pixels binarised divided by the total pixels binarised. It is given as follows:

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \tag{8}$$

where *TruePositive* means the set of pixels that are foreground in both ground truth (GT) and the binarised image. *FalsePositive* denotes the set of pixels identified as foreground in the binarised image but are background in the ground truth image.

Precision is crucial because, in document image binarisation, misidentifying background as text (false positive) can lead to erroneous or unnecessary text segmentation.

4.1.2 Recall

Recall (or sensitivity) value represents the true pixels binarised divided by the total number of true pixels. It is given as follows:

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \tag{9}$$

where *TruePositive* means the set of pixels that are foreground in both ground truth and binarised image. *FalseNegative* denotes the set of pixels identified as background in binarised image but they are foreground in ground truth image.

Recall is essential to ensure that all text (especially fine details or smaller characters) is captured.

4.1.3 F-measure

F-measure is the harmonic measure of precision and recall. Its value should be high for better results.

$$F_\beta = (1 + \beta^2) * \frac{Precision \times Recall}{\beta^2 * Precision + Recall} \tag{10}$$

With $\beta = 1$, the standard F-score is obtained as follows:

$$F_1 = \frac{2 * Precision * Recall}{(Precision + Recall)} \tag{11}$$

F1-measure is an overall measure of binarisation performance and is particularly useful for handling document images where there might be unbalanced areas of text and background (heavy inked text versus blank space, for instance).

4.1.4 Pseudo F-measure

The pseudo F-measure formula is the same of that F-measure with the exception that the ground truth is skeletonised and the pixels in the skeleton are used as foreground pixels in the calculation of pseudo F-measure.

4.1.5 PSNR

Peak signal noise ratio (PSNR) is used to check the similarity between two images, generally the constructed image and the original one. It is used for images having noise. PSNR is given as follows:

$$PSNR = 10 \log\left(\frac{C^2}{MSE}\right) \tag{12}$$

where

$$MSE = \frac{\sum_{x=1}^{M} \sum_{y=1}^{N} (I(x, y) - I'(x, y))^2}{MN} \tag{13}$$

C is a constant and mean square error (MSE) depicts the difference between the distorted image and the original one. PSNR is a measure of how close is an image to another. The MSE between the reference image and the binarised one can capture how well the binarisation method preserves the original details and structure, and the PSNR value quantifies this preservation. The higher the value of PSNR is, the higher the similarity between the images is.

4.1.6 DRD

The distance reciprocal distortion metric (DRD) has been used to measure the visual distortion in binary document images (Lu et al., 2004). It properly correlates with the human visual perception and it measures the distortion for all the S flipped pixels as follows:

$$DRD = \frac{\sum_{k=1}^{S} DRD_k}{NUBN} \tag{14}$$

where NUBN is the number of the non-uniform (not all black or white pixels) 8×8 blocks in the GT image, and DRD_k is the distortion of the kth flipped pixel that is calculated using a 5×5 normalised weight matrix W_{Nm} as defined in Sari et al. (2012). DRD_k equals to the weighted sum of the pixels in the 5×5 block of the GT that differ from the centred k^{th} flipped pixel at (x, y) in the binarisation result image B.

$$DRD_k = \sum_{i=-2}^{2} \sum_{j=-2}^{2} \left| GT_k(i, j) - B_k(x, y) \right| \times W_{Nm}(i, j) \tag{15}$$

Lower DRD values indicate that the binarisation process has introduced less distortion, meaning that the original document content is well-preserved during binarisation. Higher DRD values indicate that significant distortion has occurred, which could mean the binarisation has removed or altered important features of the document, such as fine text details, edges, or background patterns.

4.1.7 Pseudo recall

The pseudo recall formula is the same of that recall with the exception that the ground truth is skeletonised and the pixels in the skeleton are used as foreground pixels in the calculation of pseudo recall.

4.1.8 NRM

Negative rate metric (NRM), which is the proportion of foreground pixels that are incorrectly classified as background and vice versa, is calculated using the unmatched pixels between the binarised image and the ground truth image. It is given as follows:

$$NRM = \frac{R_{FN} + R_{FP}}{2} \tag{16}$$

where

$$R_{FN} = \frac{FN}{FN + TP} \quad \text{and} \quad R_{FP} = \frac{FP}{FP + TN}$$

R_{FN} is the rate of false negatives and R_{FP} is the rate of false positives. Lesser the value of NRM is, better the result of binarisation is.

4.1.9 MPM

Misclassification penalty metric (MPM) is used to evaluate how the binarised image constitutes the contour of the ground truth image. It is given as follows:

$$MPM = \frac{FP_{FN} + MP_{FP}}{2} \tag{17}$$

where

$$MP_{FN} = \frac{\sum_{i=1}^{N_{FN}} d_{FN}^{i}}{D}, \quad \text{and} \quad MP_{FP} = \frac{\sum_{j=1}^{N_{FP}} d_{FP}^{j}}{D}$$

d_{FN}^{i} and d_{FP}^{i} represent respectively the ith false negative and false positive pixel. Normalisation factor is given by D, the sum of overall pixel-to-contour distances of the GT object. Smaller the value of MPM better is the quality of the algorithm.

4.1.10   Accuracy

Accuracy-based metric is based on fundamental metrics to assess the performance of binarisation in terms of its classification of pixels as foreground (text or document content) and background (non-text).

$$Accuracy = \frac{TruePositive + TrueNegative}{TruePositive + TrueNegative + FalsePositive + FalseNegative} \qquad (18)$$

4.1.11   Specificity

Specificity is another important metric used to evaluate binarisation techniques, particularly in the context of document image binarisation. Specificity refers to the ability of a binarisation method to correctly identify background pixels as background (not text or foreground). It helps to assess how well the binarisation technique avoids false positives (mistaking background areas as text or important content). Mathematically, specificity is expressed as follows:

$$Specificity = \frac{TrueNegative}{TrueNegative + FalsePositive} \qquad (19)$$

4.1.12   Mis-classification error

Mis-classification error (ME) refers to the error that occurs when pixels in a greyscale image are incorrectly classified as either background (white) or foreground (black).

$$ME = \frac{No.\ of\ incorrectly\ predicted\ samples}{Total\ number\ of\ images} * 100 \qquad (20)$$

4.1.13   Cohen's kappa

Cohen's kappa (Congalton, 1991) compares the observed accuracy with an expected accuracy. This coefficient has an excellent correspondence with the image-quality perception by human visual inspection of the resulting images (Lins et al., 2021).

$$k = \frac{P_O - P_C}{1 - P_C}$$

P_O is the number of correctly mapped pixels (accuracy) and P_C is calculated by:

$$P_C = \frac{n_{bf} * n_{gf} + n_{bb} * n_{gb}}{N^2}$$

where n_{bf} and n_{bb} are the number of pixels mapped as foreground and background on the binary image, respectively, while n_{gf} and n_{gb} are the number of foreground and background pixels on the ground truth image and N is the total number of pixels.

4.2   *Datasets*

As depicted in Tables 1 and 2, we consider here 12 datasets the more considered in the literature. Figure 27 gives some images from DIBCO'13 as an example. For more information about the considered document image datasets, readers should refer to Nikolaidou et al. (2022) and Mustafa et al. (2019b). It is worthy to note that there are other used datasets (not addressed here) we do not consider as more frequently used like: PHIBD (Ayatollahi and Nafchi, 2013), Synchromedia Multispectral Ancient Document Images Dataset 'SMADI' (Hedjam and Cheriet, 2013), Bickley Diary (Deng et al., 2010), WEZUT OCR dataset (Michalak and Okarma, 2020), mobile-dataset-4 (Lins et al., 2022), Rahul Sharma dataset-1 (Rani et al., 2023), Nabuco dataset (Lins, 2011), Nabuco competition (Lins, 2011), LiveMemory (Alves et al., 2011), PRImA (Clausner et al., 2015), MoDI dataset (Guo et al., 2021), LRDE DBD dataset (Suh et al., 2022), e-DIBCO and e-LRDE DBD dataset with the addition of moiré patterns (Guo et al., 2021), shipping label image dataset (Suh et al., 2022), CMATERdb 6 dataset (Mollah et al., 2012). For more information, readers can ask (Ju et al., 2024; Bloechle et al., 2024; Rani et al., 2023; Lins, 2011; Alves et al., 2011; Clausner et al., 2015; Guo et al., 2021; Suh et al., 2021; Mollah et al., 2012), respectively.

Table 1   the considered datasets.

Image dataset	Number of images	Its website to be downloaded
HDIBCO2009	10	http://www.iit.demokritos.gr/~bgat/DIBCO2009/benchmark
HDIBCO2010	10	http://www.iit.demokritos.gr/~bgat/H-DIBCO2010/benchmark
DIBCO2011	16	http://utopia.duth.gr/~ipratika/DIBCO2011/benchmark
H-DIBCO2012	14	http://utopia.duth.gr/~ipratika/HDIBCO2012/benchmark
DIBCO2013	16	http://utopia.duth.gr/~ipratika/DIBCO2013/benchmark
HDIBCO2014	10	http://users.iit.demokritos.gr/~bgat/HDIBCO2014/benchmark
DIBCO2016	10	http://vc.ee.duth.gr/h-dibco2016/benchmark
DIBCO2017	20	http://vc.ee.duth.gr/dibco2017/benchmark/
DIBCO2018	10	http://vc.ee.duth.gr/h-dibco2018/benchmark/
DIBCO2019	20	http://vc.ee.duth.gr/dibco2019/benchmar/
Synthesis Database	150	https://drive.google.com/drive/folders/15WrU2EnM06wj2AJ4 9EgOLHVb1yqhPW5a?usp=sharing
Harvard Dataset	813	http://ocp.hul.harvard.edu/ihp/manuscripts.html

Table 2   The different degradations of the considered datasets

| Image dataset		Characteristics in terms of degradation
HDIBCO2009	1	Images suffer from different levels of noise such as salt-and-pepper noise and Gaussian noise.
	2	Variations in lighting during scanning lead to shadows and uneven brightness.
	3	Many images have low contrast between the text and background, making binarisation more challenging.
	4	Text from the reverse side of the page might be visible, making the separation of foreground text and background more difficult (bleed-through).
	5	The text can be faded or faint adding complexity to binarisation.
HDIBCO2010	1	Images may have a higher level of noise compared to DIBCO 2009, further complicating the binarisation task.
	2	Text may be more distorted due to factors like improper scanning or physical document damage.
	3	Some images may have a skew or rotation.
	4	The background of images could be more intricate or patterned, making it harder to distinguish text from the background.
	5	The bleed-through effect is more pronounced in some images, making it more challenging to extract the correct text.
DIBCO2011	1	The ink from the front page can bleed through to the back page, causing significant interference when trying to isolate the text from the background.
	2	Some images may have faded or discoloured text due aging or scanning issues, making it hard to distinguish between the text and background.
	3	Document text may exhibit significant warping or distortion due to folding, crinkling, or poor quality during scanning.
	4	Images have more complex backgrounds, such as noise pattern or inconsistent textures that further complicate the binarisation process.
	5	Images have non-uniform layouts. The text in these documents might not follow a clean and straight line making binarisation harder.
H-DIBCO2012	1	Images include a variety of layouts. They may contain multi-column text, often with irregular spacing or alignment, mixed-font styles, tables, figures, and illustrations.
	2	Images contain high levels of noise such as salt-and-pepper and Gaussian noise, fading, ink bleed-through, stains, and damage.
	3	The text might not follow a straight line due to the document being poorly aligned during scanning.
	4	In some images, the text from adjacent columns or lines might overlap, adding complexity to the binarisation task.

Table 2   The different degradations of the considered datasets (continued)

Image dataset		Characteristics in terms of degradation
DIBCO2013	1	Text is significantly faded, making it extremely difficult to distinguish between text and background, especially in regions with very low contrast.
	2	Ink from the opposite side of the page bleeds through, interfering with the text and complicating the binarisation process.
	3	Documents are often physically damaged, with stains, tears, and creases visible in the images.
	4	Documents may show signs of straining, such as coffee or ink stains, which can confuse binarisation algorithms.
HDIBCO2014	1	More subtle noise patterns compared to previous datasets.
	2	Images contain more intricate backgrounds that interfere with text extraction.
	3	Continued fading of the text, but not as severe as DIBCO 2011.
	4	Bleed-through stills present, but less severe than previous years.
	5	Some images with skewed, misaligned, or distorted text.
DIBCO2016	1	High-resolution images with detailed backgrounds and intricate text structures.
	2	Moderate levels of salt-and-pepper noise and Gaussian noise.
	3	Significant but not extreme fading or bleed-through.
	4	Minor stains, folds, and creases.
	5	Misaligned or skewed text due to document age and scanning.
	6	Some images have multi-column layouts that complication binarisation.
DIBCO2017	1	Increased salt-and-pepper noise and pixel distortion.
	2	Overlapping text caused by poor scanning or multiple columns.
	3	Both ink fading and bleed-through are significant.
	4	Pages show signs of tears, creases, and water stains.
	5	Text misalignment and slight skewing.
	6	Text and background have minimal contrast.
DIBCO2018	1	Salt-and-pepper, Gaussian noise, and additional random noise.
	2	Significant text fading and colour discoloration.
	3	Tears, creases, and stains
	4	Noticeable bleed-through.
	5	Text distortion.
.DIBCO2019	1	High levels of noise, including salt-and-pepper and Gaussian noise.
	2	Overlapping text caused by poor alignment or scanning artefacts.
	3	Creases, stains, and tears affecting the documents.
	4	Text is heavily faded.
	5	Strong bleed-through.
	6	Complex multi-column layouts.

Table 2 The different degradations of the considered datasets (continued)

Image dataset		Characteristics in terms of degradation
Synthesis Database	1	Salt-and-pepper noise as well as Gaussian noise.
	2	Ink fading.
	3	Blur.
	4	Distortion, stains, smudges, tears, rips, and bleed-through.
Harvard Dataset	1	Fading of text is a prominent feature due to the age of the documents.
	2	The ink is often faded.
	3	The contrast between text and background is minimal.
	4	The ink from the back side of the page bleeds through the paper, making the text on the reverse side visible on the front side.
	5	Tears, creases, folds, and rips in the pages are common.
	6	Stain and discoloration due to aging.

Figure 27 Some images from DIBCO'13 and their associated ground truth (see online version for colours)
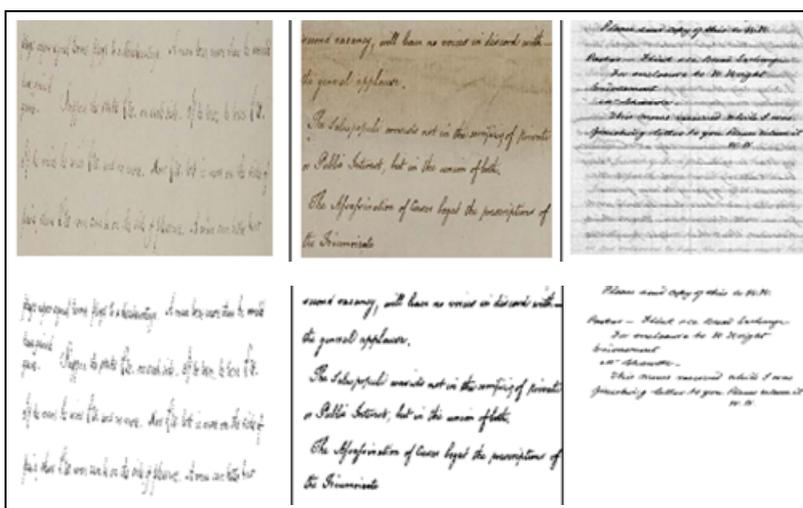
5 Experimental results and discussion

In this section, the experimental results and some associated discussions are given.

5.1 Experimental results

In this subsection, we give the results given in the different considered references. Table 3 yields the different considered datasets (12 datasets) and the associated references that use these datasets. Tables 4 and 5 provide respectively the best algorithm of the literature, without and with learning, over the different datasets and the various metrics.

Table 3   The considered references over the considered datasets

Dataset	References	Number of references
DIBCO 2009	Krupinski et al. (2019), Moghaddam and Cheriet (2010, 2012), Lu and Tan (2010), Su et al. (2010), Su et al. (2012), Gatos et al. (2009), Shi et al. (2012), Howe (2013), Su et al. (2011), Jia et al. (2018), Lu et al. (2018), Sehad et al. (2019), Xiong et al. (2021a, 2021b), Kang et al. (2021), Rani et al. (2023) and Guo et al. (2021)	18
DIBCO 2010	Krupinski et al. (2019), Su et al. (2012), Pratikakis et al. (2010), Howe (2013), Jia et al. (2018), Lu et al. (2018), Sehad et al. (2019), Xiong et al. (2021a, 2021b), Kang et al. (2021), Rani et al. (2023) and Guo et al. (2021)	12
DIBCO 2011	Krupinski et al. (2019), Su et al. (2012), Howe (2013), Jia et al. (2018), Lu et al. (2018), Vo et al. (2018), Sehad et al. (2019), Xiong et al. (2021a, 2021b), Kang et al. (2021), Ju et al. (2024), Rani et al. (2023), Guo et al. (2021) and Suh et al. (2022)	14
DIBCO 2012	Krupinski et al. (2019), Sari et al. (2014), Pratikakis et al. (2012), Jia et al. (2018), Lu et al. (2018), Sehad et al. (2019), Xiong et al. (2021a, 2021b), Kang et al. (2021), Rani et al. (2023) and Guo et al. (2021)	12
DIBCO 2013	Pratikakis et al. (2018), Jia et al. (2018), Vo et al. (2018), Sehad et al. (2019), Xiong et al. (2021a, 2021b), Kang et al. (2021), Ju et al. (2024), Sukesh et al. (2022), Guo et al. (2021) and Suh et al. (2022)	11
DIBCO 2014	Krupinski et al. (2019), Ntirogiannis et al. (2014), Jia et al. (2018), Vo et al. (2018), Xiong et al. (2021a, 2021b), Kang et al. (2021), Ju et al. (2024), Guo et al. (2021) and Suh et al. (2022)	10
DIBCO 2016	Krupinski et al. (2019), Pratikakis et al. (2016), Jana et al. (2017a, 2017b), Huang et al. (2020), Mustafa et al. (2018), Jia et al. (2018), Vo et al. (2018), Mustafa and Kader (2018b), Xiong et al. (2021a, 2021b), Kang et al. (2021), Ju et al. (2024), Guo et al. (2021) and Suh et al. (2022)	14
DIBCO 2017	Krupinski et al. (2019), Pratikakis et al. (2018), Huang et al. (2020), Xiong et al. (2021a, 2021b), Kang et al. (2021), Ju et al. (2024) and Sukesh et al. (2022)	8
DIBCO 2018	Pratikakis et al. (2018), Huang et al. (2020), Xiong et al. (2021a, 2021b), Kang et al. (2021), Ju et al. (2024) and Sukesh et al. (2022)	7
DIBCO 2019	Pratikakis et al. (2019), Xiong et al. (2021a) and Sukesh et al. (2022)	3
Synthesis	Sari et al. (2014)	1
Harvard dataset	Abd Elfattah et al. (2015)	1

Table 4 The best algorithms, with no learning, of the literature over the different datasets and the different metrics

Dataset	The best algorithm	The reference	The considered metric	The value of the considered metric	Number of considered methods in the considered reference
DIBCO 2019	Non-parametric binarisation method based on ensemble of clustering algorithms (10b in Pratikakis et al., 2019)	Bera et al. (2021) Considered in Pratikakis et al. (2019)	F-measure PF-measure PSNR DRD	72.875 72.15 14.475 16.235	24
DIBCO 2018	Combining between detecting text with stroke width transform and automatic parameter tuning (1 in Pratikakis et al., 2018)	Epstein et al. (2010) and Howe (2013) Considered in Pratikakis et al. (2018)	F-measure PSNR DRD	90.24 19.11 04.92	10
DIBCO 2017	10 in Pratikakis et al. (2018)	Considered in Pratikakis et al. (2018)	F-measure PF-measure PSNR DRD	91.04 92.86 18.28 03.40	03
DIBCO 2016	Structural symmetry of strokes	Jia et al. (2018)	F-measure PSNR MPM DRD	90.48 19.30 06.90 03.97	08
DIBCO 2014	Enhanced binarisation framework	Xiong et al. (2021b)	PF-measure	93.56	3
DIBCO 2014	Combining of Mesquita et al. (2014) and Howe (2013) (6 in Ntirogiannis et al., 2014)	Ntirogiannis et al. (2014)	F-measure PSNR DRD	96.88 22.66 0.902	10
	Enhanced binarisation framework	Xiong et al. (2021b)	PF-measure	97.73	3

Table 4   The best algorithms, with no learning, of the literature over the different datasets and the different metrics (continued)

Dataset	The best algorithm	The reference	The considered metric	The value of the considered metric	Number of considered methods in the considered reference
DIBCO 2013	Structural symmetry of strokes	Jia et al. (2018)	PF-measure	96.05	10
	Wolf	Wolf et al. (2002) Considered in Sehad et al. (2019)	DRD	01.05	6
DIBCO 2012	Enhanced binarisation framework	Xiong et al. (2021b)	F-measure PF-measure	94.26 95.16	3
	Wolf	Wolf et al. (2002) Considered in Sehad et al. (2019)	DRD	00.95	6
DIBCO 2011	Oracle	Considered in Howe (2013)	F-measure	97.50	5
	Structural symmetry of strokes	Jia et al. (2018)	PF-measure	95.09	10
	Lelore	Lelore and Bouchara (2013) Considered in Jia et al. (2018)	PSNR	19.37	
	Wolf	Wolf et al. (2002) Considered in Sehad et al. (2019)	DRD	02.42	6

Table 4 The best algorithms, with no learning, of the literature over the different datasets and the different metrics (continued)

Dataset	The best algorithm	The reference	The considered metric	The value of the considered metric	Number of considered methods in the considered reference
DIBCO 2010	Oracle	Considered in Howe (2013)	F-measure	96.50	5
	Numerated by 2: adaptive binarisation with removal of false objects and accurate local region-based active contour	Considered in Pratikakis et al. (2010)	PF-measure	95.15	17
	Enhanced binarisation framework	Xiong et al. (2021b)	PSNR	20.97	3
	Howe	Howe (2011) Considered in Jia et al. (2018)	DRD	01.72	9
DIBCO 2009	Howe	Howe (2011) Considered in Jia et al. (2018)	F-measure	94.04	10
	Lelore	Lelore and Bouchara (2013) considered in Jia et al. (2018)	PF-measure	95.10	
	Howe	Howe (2011) Considered in Jia et al. (2018)	PSNR	20.43	
			DRD	02.10	
DIBCO (2009 + 2010 + 2011 + 2012 + 2013 + 2014 + 2016)	Enhanced binarisation framework	Xiong et al. (2021b)	F-measure	93.15	3
			PSNR	20.57	

Table 5   The best algorithms, based on learning, of the literature over the different datasets and the different metrics

Dataset	The best algorithm	The reference	The considered metric	The value of the considered metric	Number of considered methods in the considered reference	Number of considered methods in our study
DIBCO 2019	DP-LinkNet	Xiong et al. (2021a)	F-measure	87.67	3	3
			PF-measure	87.56		
			PSNR	18.63		
			DRD	02.38		
DIBCO 2018	DP-LinkNet	Xiong et al. (2021a)	F-measure	95.99	3	6
			PSNR	22.71		
			DRD	01.09		
	Three-stage network using GANs	Ju et al. (2024)	PF-measure	96.96	5	
DIBCO 2017	DP-LinkNet	Xiong et al. (2021a)	F-measure	95.52	3	6
			PF-measure	96.46		
			PSNR	20.83		
			DRD	01.31		
DIBCO 2016	DP-LinkNet	Xiong et al. (2021a)	F-measure	96.29	3	7
			PF-measure	97.03		
			PSNR	23.04		
			DRD	01.05		
DIBCO 2014	DP-LinkNet	Xiong et al. (2021a)	F-measure	97.47	3	6
			PSNR	23.46		
			DRD	00.66		
	Three-stage network using GANs	Ju et al. (2024)	PF-measure	98.33	8	

Table 5 The best algorithms, based on learning, of the literature over the different datasets and the different metrics (continued)

Dataset	The best algorithm	The reference	The considered metric	The value of the considered metric	Number of considered methods in the considered reference	Number of considered methods in our study
DIBCO 2013	DP-LinkNet	Xiong et al. (2021a)	F-measure	97.15	3	7
			PF-measure	97.77		
			PSNR	24.09		
			DRD	00.78		
DIBCO 2012	DP-LinkNet	Xiong et al. (2021a)	F-measure	96.90	3	4
			PSNR	23.99		
			DRD	00.84		
DIBCO 2011	DP-LinkNet	Xiong et al. (2021a)	F-measure	96.27	3	5
			PSNR	22.23		
			DRD	01.01		
DIBCO 2010	DP-LinkNet	Xiong et al. (2021a)	F-measure	96.19	3	4
			PSNR	22.95		
			NRM	01.29		
DIBCO 2009	CMU-Nets	Kang et al. (2021)	F-measure	96.67	5	4
	DP-LinkNet	Xiong et al. (2021a)	PSNR	22.16	3	
			NRM	01.30		
DIBCO (2009 + 2010 + 2011 + 2012 + 2013 + 2014 + 2016)	DP-LinkNet	Xiong et al. (2021a)	F-measure	96.23	3	4
			PSNR	23.13		
Harvard dataset	Artificial bee colony optimiser (ABCO)	Considered in Abd Elfattah et al. (2015)	PSNR	27.20	4	1
			NRM	0.002		
			DRD	0.240		
			MPM	0.300		

5.2 Discussions

According to Tables 3, 4 and 5 based on the results given in the considered references, we can draw the following conclusions:

- There are two kinds of references: a reference as a competition where a large number of methods of the literature are experimented and a reference where a novel algorithm is introduced and is compared with some limited number of methods of the literature.

- *DIBCO 2009* dataset (the oldest database) is the most considered one among all the used databases with 18 references.

- The greatest number of methods utilised in the considered references is 24 methods experimented in Pratikakis et al. (2019) on *DIBCO 2019* followed respectively by Vo et al. (2018) and Pratikakis et al. (2010) with 18 and 17 methods for each reference.

- Broadly, there are four categories of methods: methods based on threshold (global and local or dynamic), methods based on clustering, methods based on machine learning, and the eventual combination of all these pre-cited approaches.

- In order to give an exhaustive comparison, we categorise the practical comparison into two categories: without learning and with learning. We do not mix the two categories because generally the approach with leaning outperforms largely the second class that of without learning.

- For the case 'with no learning', it seems that there is no method which outperforms really the others. However, it seems that *enhanced binarisation framework* introduced in Xiong et al. (2021b) seems to be the best over the majority of the considered datasets and especially when we consider seven datasets at once. *structural symmetry of strokes* introduced in Jia et al. (2018) may be ranked in the second position especially when on HDIBCO 2016 dataset.

- For the case 'with learning', it is clear that 'DP-LinkNet' introduced in Xiong et al. (2021a) outperforms the other considered methods.

- Broadly, methods based on deep learning comes in the first rank in terms of performance, followed by clustering while methods based on threshold comes lastly after the combination mechanism.

- Unfortunately, there is no addressing of efficiency by the majority of works dealing with binarisation. Indeed, only some ones which address efficiency like Xiong et al. (2021a). To note that we focus in this our comparison on effectiveness while efficiency goes beyond the scope of our work. Even that, it is obvious that deep learning approach, requiring training, needs more time followed by the clustering approach and dynamic threshold. Combining scheme requires also more time but it is not evident to rank it regarding to other approaches. Indeed, the theoretical ranking of combination approach, in terms of efficiency, will be clear when we know to which approaches the methods to be combined belong.

6 Conclusions

In this paper, we have given an experimental comparison study of the different techniques of document image binarisation in terms of effectiveness considering a large number of metrics and datasets. In addition, an overview of the various approaches, metrics, types of noises, and the considered datasets, is given. According to the comparative study, it is clear that the machine learning approach (or neural network approach) seems globally to be the best approach although the consuming time of learning required. As a future work, we plan to conduct another comparison study in terms of efficiency which goes beyond the scope of this paper and we plan to take into consideration more datasets and more works from the literature.

References

Abd Elfattah, M., Hassanien, A.E., Mostafa, A., Ali, A.F., Amin, K.M. and Mohamed, S. (2015) 'Artificial bee colony optimizer for historical Arabic manuscript images binarization', *2015 11th International Computer Engineering Conference (ICENCO)*, December, IEEE, pp.251–255.

Akbari, Y., AI-Maadeed, S. and Adam, K. (2020) 'Binarization of degraded document images using convolutional neural networks and wavelet-based multichannel images', *IEEE Access*, Vole. 8, No. 1, pp.153517–153534.

Almeida, M., Lines, R.D., Bernadino, R., Jesus, D. and Lima, B. (2018) 'A new binarization algorithm for historical documents', *Journal of Imaging*, Vol. 4, No. 2, pp.1–27.

Alves, N., Lencastre, M. and Lins, R.D. (2011) 'Building digital libraries of proceedings with the LiveMemory platform', *9th International Workshop on Graphics Recognition (GREC11)*, Seoul.

Aurich, V. and Weule, J. (1995) 'Non-linear Gaussian filters performing edge preserving diffusion', *Musterekennung 1995: Verstehen akustischer und visueller Informationen*, Springer Berlin Heidelberg, pp.538–545.

Ayatollahi, S.M. and Nafchi, H.Z. (2013) 'Persian heritage image binarization competition (PHIBC 2012)', *2013 First Iranian Conference on Pattern Recognition and Image Analysis (PRIA)*, March, IEEE, pp.1–4.

Badekas, E. and Papamarkos, N. (2007) 'Optimal combination of document binarization techniques using a self-organizing map neural network', *Engineering Applications of Artificial Intelligence*, Vol. 20, No. 1, pp.11–24.

Bera, S.K., Ghosh, S., Bhowmik, S., Sarkar, R. and Nasipuri, M. (2021) 'A non-parametric binarization method based on ensemble of clustering algorithms', *Multimedia Tools and Applications*, Vol. 80, No. 1, pp.7653–7673.

Bernsen, J. (1986) 'Dynamic thresholding of gray-level images', *Proceedings of 8th International Conference on Pattern Recognition*.

Bhowmik, S. (2023) 'Document image binarization', *Document Layout Analysis*, pp.11–30, Springer Nature Singapore, Singapore.

Biswas, B., Bhattacharya, U. and Chaudhuri, B.B. (2014) 'A global-to-local approach to binarization of degraded document images', *2014 22nd International Conference on Pattern Recognition*, August, IEEE, pp.3008–3013.

Bloechle, J.L., Hennebert, J. and Gisler, C. (2024) 'ZigZag: a robust adaptive approach to non-uniformly illuminated document image binarization', *Proceedings of the ACM Symposium on Document Engineering 2024*, August, pp.1–10.

Bradley, D. and Roth, G. (2007) 'Adaptive thresholding using the integral image', *Journal of Graphics Tools*, Vol. 12, No. 2, pp.13–21.

Chaki, N., Shaikh, S.H., Saeed, K., Chaki, N., Shaikh, S.H. and Saeed, K. (2014) *A Comprehensive Survey on Image Binarization Techniques*, pp.5–15, Springer, India.

Chang, C.I. and Chang, C.I. (2017) 'Discrete-time Kalman filtering for hyperspectral processing', *Real-Time Recursive Hyperspectral Sample and Band Processing: Algorithm Architecture and Implementation*, pp.49–71.

Chauhan, S., Sharma, E. and Doegar, A. (2016) 'Binarization techniques for degraded document images – a review', *2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, September, IEEE, pp.163–166.

Chen, Y. and Wang, L. (2017) 'Broken and degraded document images binarization', *Neurocomputing*, Vol. 237, No. 1, pp.272–280.

Clausner, C., Papadopoulos, C., Pletschacher, S. and Antonacopoulos, A. (2015) 'The ENP image and ground truth dataset of historical newspapers', *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, August, IEEE, pp.931–935.

Congalton, R.G. (1991) 'A review of assessing the accuracy of classifications of remotely sensed data', *Remote Sensing of Environment*, Vol. 37, No. 1, pp.35–46.

Deng, F., Wu, Z., Lu, Z. and Brown, M.S. (2010) 'Binarizationshop: a user-assisted software suite for converting old documents to black-and-white', *Proceedings of the 10th Annual Joint Conference on Digital Libraries*, June, pp.255–258.

Drira, F. (2006) 'Towards restoring historic documents degraded over time', *Second International Conference on Document Image Analysis for Libraries (DIAL'06)*, April, IEEE, pp.8–357.

Epstein, B., Ofek, E. and Wexler, Y. (2010) 'Detecting text in natural scences with stroke width transform', *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June, IEEE, pp.2963–2970.

Gatos, B., Ntirogiannis, K. and Pratikakis, I. (2009) 'ICDAR 2009 document image binarization contest (DIBCO 2009)', *2009 10th International Conference on Document Analysis and Recognition*, July, IEEE, pp.1375–1382.

Gatos, B., Pratikakis, I. and Perantonis, S.J. (2006) 'Adaptive degraded document image binarization', *Patter Recognition*, Vol. 39, No. 3, pp.317–327.

Guo, Y., Ji, C., Zheng, X., Wang, Q. and Luo, X. (2021) 'Multi-scale multi-attention network for moiré document image binarization', *Signal Processing: Image Communication*, Vol. 90, No. 1, pp.1–7.

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

He, K., Sun, J. and Tang, X. (2012) 'Guided image filtering', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 6, pp.1397–1409.

Hedjam, R. and Cheriet, M. (2013) 'Historical document image restoration using multispectral imaging system', *Pattern Recognition*, Vol. 46, No. 8, pp.2297–2312.

Howe, N.R. (2011) 'A laplacian energy for document binarization', *2011 International Conference on Document Analysis and Recognition*, September, IEEE, pp.6–10.

Howe, N.R. (2013) 'Document binarization with automatic parameter tuning', *International Journal on Document Analysis and Recognition (IJDAR)*, Vol. 16, No. 3, pp.247–258.

Huang, X., Li, L., Liu, R., Xu, C. and Ye, M. (2020) 'Binarization of degraded document images with global-local U-Nets', *Optik*, Vol. 203, No. 1, pp.1–7.

Jain, A.K. (1989) *Fundamentals of Digital Image Processing*, Prentice-Hall, Inc.

Jana, P., Ghosh, S., Bera, S.K. and Sarkar, R. (2017a) 'Handwritten document image binarization: an adaptive k-means based approach', *2017 IEEE Cacutta Conference (CALCON)*, December, IEEE, pp.226–230.

Jana, P., Ghosh, S., Sarkar, R. and Nasipuri, M. (2017b) 'A fuzzy C-means based approach towards efficient document image binarization', *2017 Ninth International Conference on Advances in Pattern Recognition (ICAPR)*, December, IEEE, pp.1–6.

Jia, F., Shi, C., He, K., Wang, C. and Xiao, B. (2018) 'Degraded document image binarization using structural symmetry of strokes', *Pattern Recognition*, Vol. 74, No. 1, pp.225–240.

Johannsen, G. (1982) 'A threshold selection method using information measures', *International Conference on Pattern Recognition (ICPR'82)*, pp.140–143.

Ju, R.Y., Lin, Y.S., Jin, Y, Chen, C.C., Chien, C.T. and Chiang, J.S. (2024) 'Three-stage binarization of color document images based on discrete wavelet transform and generative adversarial networks', *Knowledge-Based Systems*, Vol. 304, No. 1, pp.1–12.

Kang, S., Iwana, B.K. and Uchida, S. (2021) 'Complex image processing with less data – document image binarization by integrating multiple pre-trained U-Net modules', *Pattern Recognition*, Vol. 109, No. 1, pp.1–14.

Kapur, J.N., Sahoo, P.K. and Wong, A.K. (1985) 'A new method for gray-level picture thresholding using the entropy of the histogram', *Computer Vision, Graphics, and Image Processing*, Vol. 29, No. 3, pp.273–285.

Kefali, A., Sari, T. and Bahi, H. (2014) 'Foreground-background separation by feed-forward neural networks in old manuscripts', *Informatica*, Vol. 38, No. 4, pp.329–338.

Khurshid, K., Siddiqi, I., Faure, C. and Vincent, N. (2009) 'Comparison of Niblack inspired binarization methods for ancient documents', *Document Recognition and Retrieval XVI*, January, Vol. 7247, No. 1, pp.267–275, SPIE.

Kim, I.J. (2004) 'Multi-window binarization of camera image for document recognition', *Ninth International Workshop on Frontiers in Handwriting Recognition*, October, IEEE, pp.323–327.

Krupinski, R., Lech, P., Teclaw, M. and Okarma, K. (2019) 'Binarization of degraded document images with generalized Gaussian distribution', *Computational Science-ICCS 2019: 19th International Conference, Proceedings, Part V*, 12–14 June, Springer International Publishing, Faro, Portugal, pp.177–190.

Kwon, S.H. (2004) 'Threshold selection based on cluster analysis', *Pattern Recognition Letters*, Vol. 25, No. 9, pp.1045–1050.

Lelore, T. and Bouchara, F. (2013) 'FAIR: a fast algorithm for document image restoration', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 8, pp.2039–2048.

Li, Y., Jin, W., Zhu, J., Zhang, X. and Li, S. (2018) 'An adaptive deghosting method in neural network-based infrared detectors nonuniformity correction', *Sensors*, Vol. 18, No. 1, pp.1–15.

Lins, R.D. (2011) 'Nabuco-two decades of processing historical documents in Latin America', *J. Univers. Comput. Sci.*, Vol. 17, No. 1, pp.151–161.

Lins, R.D., Bernardino, R.B., Barboza, R. and Oliveira, R. (2022) 'The winner takes it all: choosing the 'best' binarization algorithm for photographed documents', *International Workshop on Document Analysis Systems*, May, Springer International Publishing, Cham, pp.48–64.

Lins, R.D., Bernardino, R.B., Smith, E.B. and Kavallieratou, E. (2021) 'ICDAR 2021 competition on time-quality document image binarization', *Document Analysis and Recognition-ICDAR 2021: 16th International Conference, Proceedings, Part IV*, 5–10 September, Springer International Publishing, Lausanne, Switzerland, pp.708–722.

Liu, Y. and Srihari, S.N. (1997) 'Document image binarization based on texture features', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 5, pp.540–544.

Lu, D., Huang, X. and Sui, L. (2018) 'Binarization of degraded document images based on contrast enhancement', *International Journal on Document Analysis and Recognition (IJDAR)*, Vol. 21, No. 1, pp.123–135.

Lu, H., Kot, A.C. and Shi, Y.Q. (2004) 'Distance-reciprocal distortion measure for binary document images', *IEEE Signal Processing Letters,* Vol. 11, No. 2, pp.228–231.

Lu, S., B. and Tan, C.L. (2010) 'Document image binarization using background estimation and stroke edges', *International Journal on Document Analysis and Recognition (IJDAR)*, Vol. 13, No. 1, pp.303–314.

Mesquita, R.G., Mello, C.A. and Almeida, L.H.E.V. (2014) 'A new thresholding images based on the perception of objects by distance', *Integrated Computer-Aided Engineering*, Vol. 21, No. 2, pp.133–146.

Michalak, H. and Okarma, K. (2019) 'Improvement of image binarization methods using image preprocessing with local entropy filtering for alphanumerical character recognition purposes', *Entropy*, Vol. 21, No. 6, pp.1–18.

Michalak, H. and Okarma, K. (2020) 'Robust combined binarization method of non-uniformly illuminated document images for alphanumerical character recognition', *Sensors*, Vol. 20, No. 10, pp.1–23.

Moghaddam, R.F. and Cheriet, M. (2010) 'A multi-scale framework for adaptive binarization of degraded document images', *Pattern Recognition*, Vol. 43, No. 6, pp.2186–2198.

Moghaddam, R.F. and Cheriet, M. (2012) 'AdOtsu: an adaptive and parameterless generalization of Otsu's method for document image binarization', *Pattern Recognition*, Vol. 45, No. 6, pp.2419–2431.

Mollah, A.F., Basu, S. and Nasipuri, M. (2012) 'Computationally efficient implementation of convolution-based locally adaptive binarization techniques', *International Conference on Information Processing*, August, Springer Berlin Heidelberg, Berlin, Heidelberg, pp.159–168.

Mondal, T., Coustaty, M., Gomez-Kramer, P. and Ogier, J.M. (2019) 'Learning free document image binarization based on fast fuzzy c-means clustering', *2019 International Conference on Document Analysis and Recognition (ICDAR)*, September, IEEE, pp.1384–1389.

Mustafa, W.A. and Kader, M.M.M.A. (2018a) 'Binarization of document images: a comprehensive review', *Journal of Physics: Conference Series*, June, IOP Publishing, Vol. 1019, No. 1, pp.1–10.

Mustafa, W.A. and Kader, M.M.M.A. (2018b) 'Binarization of document image using optimum threshold modification', *Journal of Physics: Conference Series*, June, IOP Publishing, Vol. 1019, No. 1, pp.1–9.

Mustafa, W.A., Aziz, H., Khairunizam, W., Ibrahim, Z., Shahriman, A.B. and Razlan, Z.M. (2018) 'Review of different binarization approaches on degraded document images', *2018 International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA)*, August, IEEE, pp.1–8.

Mustafa, W.A., Aziz, H., Khairunizam, W., Zunaidi, I., Razlan, Z.M. and Shahriman, A.B. (2019a) 'Document images binarization using hybrid combination of fuzzy C-means and deghost method', *IOP Conference Series: Materials Science and Engineering*, June, IOP Publishing, Vol. 557, No. 1, pp.1–10.

Mustafa, W.A., Khairunizam, W., Zunaidi, I., Razlan, Z.M. and Shahriman, A.B. (2019b) 'A comprehensive review on document image (DIBCO) database', *IOP Conference Series: Materials Science and Engineering*, June, IOP Publishing, Vol. 557, No. 1, pp.1–9.

Nafchi, H.Z., Moghaddam, R.F. and Cheriet, M. (2013) 'Historical document binarization based on phase information of images', *Computer Vision-ACCV 2012 Workshops: ACCV 2012 International Workshops, Revised Selected Papers, Part II*, 5–6 November, Springer Berlin Heidelberg, Daejeon, Korea, pp.1–12.

Niblack, W. (1985) *An Introduction to Digital Image Processing*, Strandberg Publishing Company, Denmark.

Nikolaidou, K., Seuret, M., Mokayed, H. and Liwicki, M. (2022) 'A survey of historical document image datasets', *International Journal on Document Analysis and Recognition (IJDAR)*, Vol. 25, No. 4, pp.305–338.

Ntirogiannis, K., Gatos, B. and Pratikakis, I. (2014) 'ICFHR2014 competition on handwritten document image binarization (H-DIBCO 2014)', *2014 14th International Conference on Frontiers in Handwriting Recognition*, September, IEEE, pp.809–813.

Otsu, N. (1979) 'A threshold selection method from gray-level histograms', *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 9, No. 1, pp.62–66.

Papamarkos, N. (2001)' A technique for fuzzy document binarization', *Proceedings of the 2001 ACM Symposium on Document Engineering*, November, pp.152–156.

Pierre, D. (1993) *WELLNER Adaptive Thresholding for the Digital Desk*, Technical Report EPC-1993-110.

Pratikakis, I., Gatos, B. and Ntirogiannis, K. (2010) 'H-DIBCO 2010-handwritten document image binarization competition', *2010 12th International Conference on Frontiers in Handwriting Recognition*, November, IEEE, pp.727–732.

Pratikakis, I., Gatos, B. and Ntirogiannis, K. (2012) 'ICFHR 2012 competition on handwritten document image binarization (H-DIBCO 2012)', *2012 International Conference on Frontiers in Handwriting Recognition*, September, IEEE, pp.817–822.

Pratikakis, I., Zagoris, K., Barlas, G. and Gatos, B. (2016)' ICFHR2016 handwritten document image binarization contest (H-DIBCO 2016)', *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, October, IEEE, pp.619–623.

Pratikakis, I., Zagoris, K., Kaddas, P. and Gatos, B. (2018) 'ICFHR2018 Competition on handwritten document image binarization contest (H-DIBCO 2018)', *International Conference on Frontiers in Handwritting Recognition*, p.1.

Pratikakis, I., Zagoris, K., Karagiannis, S., Tsochatzidis, L., Mondal, T., Wang, X, … and Li, M. (2019) *ICDAR 2019 Competition on Document Image Binarization*, November.

Rahiche, A. and Cheriet, M. (2019) 'KFBin alman filter-based approach for document image binarization', *Image Analysis and Recognition: 16th International Conference, ICIAR 2019, Proceedings, Part I*, 27–29 August, Springer International Publishing, Waterloo, ON, Canada, pp.150–161.

Rani, U., Kaur, A. and Josan, G. (2023) 'A new binarization method for degraded document images', *International Journal of Information Technology*, Vol. 15, No. 2, pp.1035–1053.

Sari, T., Kefali, A. and Bahi, H. (2012) 'An MLP for binarizing images of old manuscripts', *2012 International Conference on Frontiers in Handwriting Recognition*, September, IEEE, pp.247–251.

Sari, T., Kefali, A. and Bahi, H. (2014) 'Text extraction from historical document images by the combination of several thresholding techniques', *Advances in Multimedia,* Vol. 2014, No. 1, pp.1–10.

Sarkka, S. and Svensson, L. (2023) *Bayesian Filtering and Smooting*, Vol. 17, Cambridge University Press, England.

Sauvola, J. and Pietikainen, M. (2000) 'Adaptive document image binarization', *Pattern recognition*, Vol. 33, No. 2, pp.225–236.

Sehad, A., Chibani, Y., Hedjam, R. and Cheriet, M. (2019) 'Gabor filter-based texture for ancient degraded document image binarization', *Pattern Analysis and Applications*, Vol. 22, No. 1, pp.1–22.

Sezgin, M. and Sankur, B.L. (2004) 'Survey over image thresholding techniques and quantitative performance evaluation', *Journal of Electronic Imaging*, Vol. 13, No. 1, pp.146–168.

Shi, J., Ray, N. and Zhang, H. (2012) 'Shape based local thresholding for binarization of document images', *Pattern Recognition Letters*, Vol. 33, No. 1, pp.24–32.

Su, B., Lu, S. and Tan, C.L. (2010) 'Binarization of historical document images using the local maximum and minimum', *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*, June, pp.159–166.

Su, B., Lu, S. and Tan, C.L. (2011) 'Combination of document image binarization techniques', *2011 International Conference on Document Analysis and Recognition*, September, IEEE, pp.22–26.

Su, B., Lu, S. and Tan, C.L. (2012) 'Robust document image binarization technique for degraded document images', *IEEE Transactions on Image Processing*, Vol. 22, No. 4, pp.1408–1417.

Suh, S., Kim, J., Lukowicz, P. and Lee, Y.O. (2022) 'Two-stage generative adversarial networks for binarization of color document images', *Pattern Recognition*, Vol. 130, No. 1, pp.1–11.

Suh, S., Lukowicz, P. and Lee, Y.O. (2021) 'Fusion of global-local features for image quality inspection of shipping label', *2020 25th International Conference on Pattern Recognition (ICPR)*, January, IEEE, pp.2643–2649.

Sukesh, R., Seuret, M., Nicolaou, A., Mayr, M. and Christlein, V. (2022)' A fair evaluation of various deep learning-based document image binarization approaches', *International Workshop on Document Analysis Systems*, May, Springer International Publishing, Cham, pp.771–785.

Sulaiman, A., Omar, K. and Nasrudin, M.F. (2019) 'Degraded historical document binarization: a review on issues, challenges, techniques, and future directions', *Journal of Imaging*, Vol. 5, No. 4, pp.1–25.

Tensmeyer, C. and Martinez, T. (2017) 'Document image binarization with fully convolutional neural networks', *2017 14th IAPR International Conference on document analysis and recognition (ICDAR)*, November, IEEE, pp.99–104.

Tensmeyer, C. and Martinez, T. (2020) 'Historical document image binarization: a review', *SN Computer Science*, Vol. 1, No. 3, p.173.

Tomasi, C. and Manduchi, R. (1998) 'Bilateral filtering for gray and color images', *Sixth International Conference on Computer Vision*, January, IEEE, IEEE Cat. No. 98CH36271, pp.839–846.

Tong, L.J., Chen, K., Zhang, Y., Fu, X.L. and Duan, J.Y. (2009) 'Document image binarization based on NFCM', *2009 2nd International Congress on Image and Signal Processing*, October, IEEE, pp.1–5.

Vo, Q.N., Kim, S.H., Yang, H.J. and Lee, G. (2018) 'Binarization of degraded document images based on hierarchical deep supervised network', *Pattern Recognition*, Vol. 74, No. 1, pp.568–586.

Wen, J., Li, S. and Sun, J. (2013) 'A new binarization method for non-uniform illuminated document images', *Pattern Recognition*, Vol. 46, No. 6, pp.1670–1690.

Wolf, C., Jolion, J.M. and Chassaing, F. (2002) 'Text localization, enhancement and binarization in multimedia documents', *2002 International Conference on Pattern Recognition*, August, IEEE, pp.1037–1040.

Woods, R.E. and Gonzalez, R.C. (2008) *Digital Image Processing*, Pearson Education, India.

Wu, Y., Natarajan, P., Rawls, S. and AbdAlmageed, W. (2016) 'Learning document image binarization from data', *2006 IIEEE International Conference on Image Processing (ICIP)*, September, IEEE, pp.3763–3767.

Xiong, W., Jia, X., Yang, D., Ai, M., Li, L. and Wang, S. (2021a) 'DP-LinkNet: a convolutional network for historical document image binarization', *KSII Transactions on Internet and Information Systems (TIIS)*, Vol. 15, No. 5, pp.1778–1797.

Xiong, W., Zhou, L., Yue, L., Li, L. and Wang, S. (2021b) 'An enhanced binarization framework for degraded historical document images', *EURASIP Journal on Image and Video Processing*, Vol. 2021, No. 1, pp.1–24.

Yanowitz, S.D. and Bruckstein, A.M. (1989) 'A new method for image segmentation', *Computer vision, Graphics, and Image Processing*, Vol. 46, No. 1, pp.82–95.

Yazid, H. and Arof, H. (2013) 'Gradient based adaptive thresholding', *Journal of Visual Communication and Image Representation*, Vol. 24, No. 7, pp.926–936.

Yoon, Y., Ban, K.D., Yoon, H., Lee, J. and Kim, J. (2013) 'Best combination of binarization methods for license plate character segmentation', *ETRI Journal*, Vol. 35, No. 3, pp.491–500.

Zhou, Y., Zuo, S., Yang, Z., He, Z., He, J., Shi, J. and Zhang, R. (2023) 'A review of document image enhancement based on document degradation problem', *Applied Sciences*, Vol. 13, No. 13, pp.1–23.