IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 20, 2020, accepted August 15, 2020, date of publication August 19, 2020, date of current version September 1, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3017783

Binarization of Degraded Document Images Using Convolutional Neural Networks and Wavelet-Based Multichannel Images

YOUNES AKBARI, SOMAYA AL-MAADEED, (Senior Member, IEEE), AND KALTHOUM ADAM, (Member, IEEE)

Department of Computer Science and Engineering, Qatar University, Doha, Qatar

Corresponding author: Somaya Al-Maadeed (s_alali@qu.edu.qa)

This work was supported by the Qatar National Research Fund (a member of Qatar Foundation) through the National Priority Research Program (NPRP) under Grant NPRP 7-442-1-082.

ABSTRACT Convolutional neural networks (CNNs) have previously been broadly utilized to binarize document images. These methods have problems when faced with degraded historical documents. This paper proposes the utilization of CNNs to identify foreground pixels using novel input-generated multichannel images. To create the images, the original source image is decomposed into wavelet subbands. Then, the original image is approximated by each subband separately, and finally, the multichannel image is constituted by arranging the original source image (grayscale image) as the first channel and the approximated image by each subband as the remaining channels. To achieve the best results, 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. To investigate the effect of the multichannel images proposed as network inputs, the CNNs used in the architectures are three popular networks, namely, U-net, SegNet, and DeepLabv3+. The experimental results of the scenarios demonstrate that our method is more successful than the three CNNs when trained by the original source images and proves competitive performance in comparison with state-of-the-art results using the DIBCO database.

INDEX TERMS Document image binarization, wavelet-based multichannel images, single and multiple CNNs, SegNet, U-net, DeepLabv3+.

I. INTRODUCTION

Binarization in document analysis aims to distinguish text as foreground pixels from the background. This task is one of the preprocessing steps that has a significant impact on steps such as feature extraction and recognition from document images that require a high-quality and accurate foreground [1]–[3]. When a handwritten document is scanned, the quality and curvature of the handwriting pose serious problems. Nevertheless, historical documents provide an additional challenge that can lead to degradation. The degradation can be increased by smudges, stains, bleed-through, and disappearance of the letters in considerable segments of a document. These issues have encouraged the document analysis community to focus on several document image binarization competitions (DIBCOs) in concurrence with conferences such as the ICFHR and ICDAR beginning in 2009 [4]. Some ideas and techniques in other fields can be used in document binarization such as semantic segmentation, image denoising, background removal, and image restoration. In recent years, the document analysis community, to achieve better performance of the binarization problem, has applied new aspects of deep learning techniques. However, certain challenges remained after using the mentioned method. For instance, two limitations affect the perfection of the function of these methods [5]. The first limitation occurs when background noise in historical documents due to aging remains after the binarization process, especially if the noise is located away from the text, and the foreground and background have the same information, which could create a false positive effect, as shown in Fig. 1(a) and (c). This challenge occurs when many writers have used one document to write their texts over the course of many years. The aim of the challenge is to detect the last text as a foreground. The second limitation

The associate editor coordinating the review of this manuscript and approving it for publication was Hongjun Su.

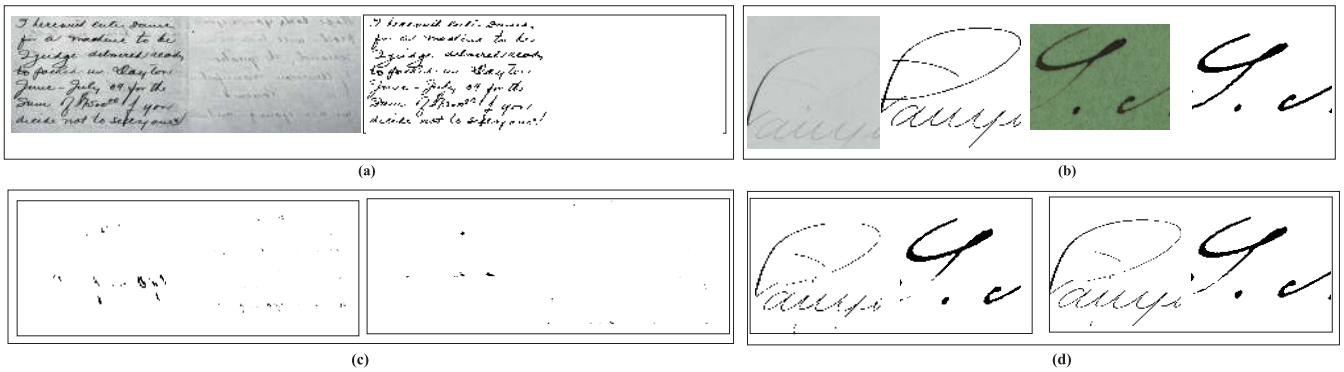

FIGURE 1. Considering the two important weaknesses in document binarization, two methods are applied to solve the problems (SegNet method based on original images and our proposed method): (a) and (b): original and ground truth images, (c) left: SegNet method, and right: our method (as shown, our method significantly improves the problem of text-like background noise). (c) left: SegNet method, and right: our method (as shown, our method is better at solving the problem of thin/blurred strokes).

occurs when a subset of text with faded strokes, as displayed in Fig. 1(b) and (d), is wrongly recognized as noise and then omitted from the foreground in the binarization procedure. As a result of the previous errors, finding new methods for increasing the accuracy of the binarized image is needed.

In addition, essential structures of the documents (such as endpoints, corners, and edges) might fail when CNNs use the original images as an input. Although methods based on CNNs exploit sophisticated machine-learning structures in the spectral domain or the original spatial domain, they do not emphasize image characteristics in the transform domain [6].

The wavelet transform [7], [8] can be applied to address the above mentioned problem by considering the wavelet coefficient in both the low- and high-frequency subbands. The values of the wavelet coefficient with respect to the low- and high-frequency subbands disclose important information related to the image structure. Both frequency subbands usually signify image edges and textures in which the high and low values of the wavelet coefficient correspond to complex edges or sharp textures and smooth edges or textures, respectively. High wavelet coefficients reflect high-average brightness regions, while small coefficients reflect low-average brightness regions. By leveraging this successful attribute of the wavelet transform, we can consider CNNs at different frequency bands for the aim of document binarization.

To add the advantageous attribute to CNNs based upon wavelets, we implement an innovative technique that considers the transform domain information together with the original image (grayscale image) as input data of CNNs, called multichannel images. Compared to the grayscale image that has one channel of data, our created image includes multichannel data; one of the channels is the grayscale image, and the remaining channels are created by wavelet analysis. The inspiration for our efforts to create wavelet-based multichannel images has been the reassuring output from the utilization of multispectral images [9] in semantic segmentation problems. Compared to our presented multichannel images that are based on wavelet subbands, multispectral images are created in a multichannel manner by a noninvasive imaging method for detecting invisible features of the human eye. The images as wavelet-based multichannel images are introduced. Additionally, our approach based on multichannel images is investigated in two different architectures to train our network (i.e., single- and multiscale- stream networks). Although our work on multichannel images is not the first attempt in the field and works such as [10] have explored it, multichannel images based on approximated outputs of wavelet transform is the first attempt. We explore the impact of the additional information fed into CNNs as input.

Fig. 2 shows an overview of the presented method. In the single stream, our network image-generated input is produced from the original image (grayscale mode) along with images approximated by all subbands (three subbands), which fully generate a final image with four channels. In the other approach, we consider multiple-stream CNNs. The input of each network includes the original image and one of the images approximated by one of the subbands, and the final image for each network has two channels. The binarization result of each CNN is integrated to obtain the final segmentation map. To exhibit the efficacy of the multichannel images based on CNNs and to enact a fair comparison, the created multichannel images were additionally trained on three different CNNs (SegNet, U-net and DeepLabv3+). This paper focuses on document binarization using CNNs and wavelet-based multichannel images. The predominant contributions of this paper are described as follows:

- *A novel approach for semantic segmentation (in particular, document binarization):* The main contribution is the proposal of a novel approach using (single and multiple) CNNs and a multichannel image.
- *Wavelet-based multichannel images*: This work is the first attempt to investigate wavelet-based multichannel images (to the best of the authors' knowledge). The basic idea behind creating the multichannel image in our work is multispectral images used in semantic segmentation.
- *New database based on multichannel images*: The wavelet-based multichannel images (three sets of images with two channels for training multiple CNNs

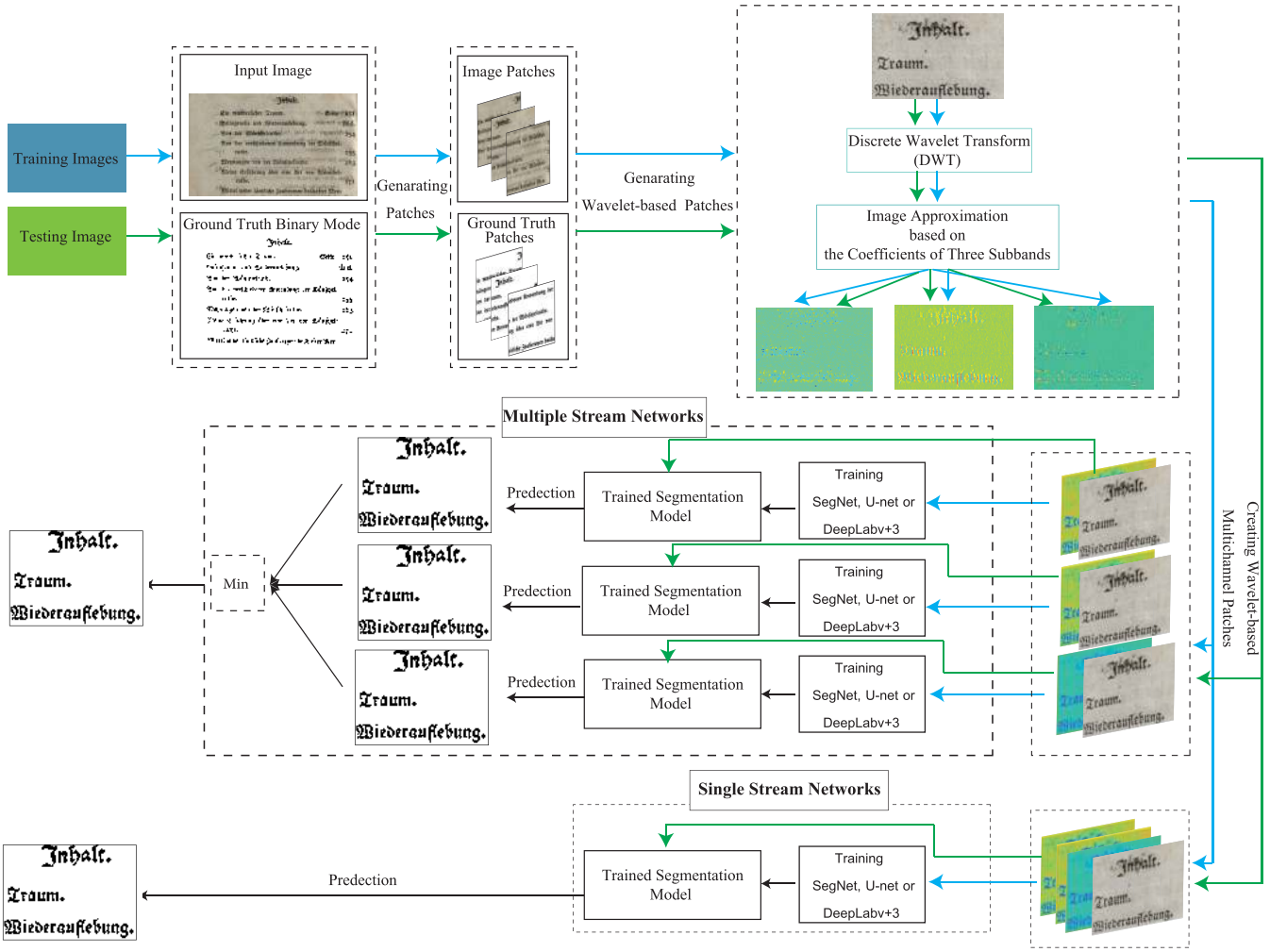

FIGURE 2. **Workflow of the proposed document binarization system (displayed wavelet subbands are based on scaled colors).**

and one set of images with four channels for single CNNs) and ground truth patches are collected to train CNNs (SegNet, U-net and DeepLabv3+) for document image binarization. The multichannel database can also be trained in other architectures.

- *Improved accuracy for document binarization*: We show that compared with the document binarization method based on SegNet, U-net and DeepLabv3+ without the use of multichannel images, the proposed methods achieve superior performance. Enhanced competitiveness and robustness of the presented method are shown in comparison with state-of-the-art results based upon four assessment measures. Finally, our method improves the two problems of text-like background noises and thin/blurred strokes.

The rest of this investigation is formulated as follows. Section II presents a literature review of related work, and section III introduces the recommended model. Experimental outputs are presented in section IV, and section V wraps up this paper.

II. RELATED WORK

In document analysis (particularly historical documents), document binarization plays a significant role in achieving better results in applications such as line, word and character segmentation and word spotting. Degradation and the poor state of historical documents complicate the binarization process of these images. Based on our review of current binarization methods, we categorize them into two categories. In the first category, we consider methods that do not use deep learning approaches, and the other category includes methods based on deep learning. In this review, we begin by presenting the two categories and subsequently discuss competitions in document binarization that have been held in conjunction with the ICFHR and ICDAR conferences. In addition, we refer readers to further details in two current and comprehensive research efforts [11], [12] in which a review of important works in the field of the document binarization has been explored and possible future research directions in this subject have been presented.

A. FIRST CATEGORY

Traditional approaches to binarization tend to classify an image based on global or local thresholds. The classification step divides the pixels into two different groups: foreground and background. For instance, the Otsu approach in [13] is based on creating the maximum separation between the background and text and depends on finding the maximum interclass variation to successfully binarize the image. Other techniques that were introduced later, such as Niblack [14], Sauvola [15], and [16], depend on the estimation of the local intensity using local threshold-based techniques. The binary image resulting from the previous techniques might consider selected background pixels as it is applied locally in each pixel neighborhood [17]. To overcome this issue, Lu *et al.* [18] suggested fusing the local and global thresholds and began by applying polynomial smoothing to classify the background and subsequently used the local threshold to capture the foreground text. Hybrid techniques use the advantages of both global and local thresholding techniques [19]. Reference [20] followed another method to increase the efficiency of the binarization process by applying different binarization methods on divided blocks of an input image. In [21], a combination of several methods was integrated to obtain a promising output. Normalization of a degraded image based on an inpainting algorithm followed by combining hybrid techniques at the connected component level was presented in [22]. One of the most important nonthreshold approaches (the Laplacian energy of the image intensity) was proposed by Howe [23]. An improved version of this method used adaptive tuning of two key parameters [24]. A new energy function introduced by Mishra *et al.* [25] iteratively used a graph-cut method to obtain optimal binarization. Degraded document images were addressed with the use of a quick adaptive thresholding method in [26]. A recently published study [27] presented an effective binarization method that obtained a threshold locally. The method is based on structural symmetric pixels (SSPs) and achieved promising results for degraded document images.

B. SECOND CATEGORY

The second category includes deep learning methods that have achieved promising results in machine vision tasks, such that the spatial dependencies and redundancy of an image are effectively encoded. One of the new deep learning applications is binarization of the degraded document image that considers a quantity of training data to train a deep neural network to assign image pixels as either background or foreground. Deep learning methods with the aim of image segmentation are designed in various types of architecture. In [28], a CNN [29] and postprocessing method based on a graph-cut was used to improve the results in [30]. The authors in [10] successfully combined an effective loss function with the Howe method [23] attributes in a single-stream network to improve CNNs. A SegNet network for semantic labeling was adapted for the document binarizations in [31], where a loss function based on the labeling uniformity and piecewise constant were considered. Peng *et al.* [32] extended the previous method by applying a fully convolutional network (FCN) encoder-decoder and a conditional random field (CRF). Westphal *et al.* [33] applied a type of recurrent neural network, named Grid-LSTM, where a pseudo F-measure was utilized for the loss function. In [34], an autoencoder for image binarization was presented based on a global threshold. The autoencoder was a simple model of the CNNs (similar to SegNet networks), that included the advantages and disadvantages of CNNs. To investigate the type and quantity of the input data, the network was trained based on specialized data and global data, including parts of databases and complete databases. The results showed that selecting specialized data does not achieve a significant improvement. A deep learning method primarily based on a CNN classic structure, namely, convolutional layers and a fully connected layer, was presented by Pastor-Pellicer *et al.* [35]. Their new novelty was a centered sliding window that was used to improve the classification step. A new formulation based on the 2D long short-term memory (LSTM) to binarize document images was introduced by Afzal *et al.* [36]. A hierarchical deep supervised network (DSN) architecture that used multiple networks with different depths was presented in [5]. The authors reported that the results outperformed the state-of-the-art methods. The method presented in [37] was improved by Ayyalasomayajula *et al.* in [28] utilizing graph cuts as a postprocessing step. [38] used images predicted by the structural symmetric pixels (SSPs) method [27] and original images to train CNNs. The method improved not only the CNNs used in their approach that were trained without a prediction step but also the results obtained by the SSPs method. A cascading modular U-net was presented in [39], in which two dilation and erosion networks were trained before the binarization network.

C. DOCUMENT IMAGE BINARIZATION CONTEST (DIBCO)

ICDAR and ICFHR succeeded in presenting most of the up-to-date developments in document image binarization. They held two international document image binarization competitions in conjunction with the two conferences. The challenges mentioned in the previous section are considered in the training and testing data (machine-printed and handwritten) of the DIBCOs, including images with color and grayscale modes, their ground truth, and the evaluation measures. In this subsection, we review the methods that achieved the best performance (winner) in the competitions. In Table 1, we list the details of the winning methods in the competitions from ICDAR 2009 to ICDAR 2019 and from ICFHR 2010 to ICFHR 2018. To assess the effectiveness of binarization methods, four suitable measures were applied in the competitions:

- *(i) F-Measure (FM)*:

$$FM = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{1}$$

TABLE 1. Detailed evaluation results (%) for winners of DIBCOs.

Competition	Winner	Method	FM (%)	F_{ps} (%)	PSNR (%)	DRD (%)
ICDAR 2009 [4]	S. Lu, C.L. Tan	Four steps: background extraction, stroke edge detection, local thresholding, and postprocessing	91.24	-	18.66	-
ICFHR 2010 [40]	B. Su and et al	Four steps: contrast evaluation, stroke edge detection, local thresholding, and postprocessing	91.50	-	19.78	-
ICDAR 2011 [41]	T. Lelore and F. Bouchara	Median filtering and upscaling with linear interpolation, text estimation and edge detection, clustering image to text, background and unknown	88.70	-	17.8	8.67
ICFHR 2012 [42]	Nicholas R. Howe	Optimizing a global energy function based on the Laplacian image [22]	89.47	-	21.80	3.44
ICDAR 2013 [43]	Bolan Su and et al	Contrast evaluation and combination with a edge map to extract an accurate text character	92.12	94.19	20.68	3.10
ICFHR 2014 [44]	R. G Mesquita and et al	Perceiving distant objects based on the human visual system proposed in [45] and combination with Howe method	020 96.88	97.65	22.66	0.90
ICFHR 2016 [46]	N. Kligler and A. Tal	Three stages: creating the visibility score map based on [47], binarization based on Howe method and post processing for reducing noises	87.61	91.28	18.11	5.21
ICDAR 2017 [48]	D. Ilin and et al	Based on U-Net convolutional network architecture [49]	91.04	92.86	18.28	3.40
ICFHR 2018 [50]	XIONG Wei and et al	Based on Howe method	88.34	90.24	19.11	4.92
ICDAR 2019 [51]	Soulib Ghosh and et al	Based on clustering algorithms	72.87	72.15	14.47	16.23

where Precision and Recall are explained in terms of true positive (TP), false positive (FP) and false negative (FN): $Precision = \frac{TP}{TP+FP}$ and $Recall = \frac{TP}{TP+FN}$.

- *(ii) pseudo-FMeasure (F_{ps})*:

$$p - FM = \frac{2 \times pRecall \times Precision}{pRecall + Precision} \tag{2}$$

where the percentage of the skeletonized ground truth image is estimated by pRecall [40], [42], [52].

- *(iii) Peak Signal to Noise Ratio (PSNR)*:

$$PSNR = 10 \log \left(\frac{C^2}{MSE} \right) \tag{3}$$

where $MSE = \frac{\sum_{M}^{x=1} \sum_{N}^{y=1} (I_{bin}(x,y) - I'_{bin}(x,y))^2}{MN}$ and C is explained as the discrepancy between the text and background. The proportion of similarity between the two images is obtained by the measure.

- *(iv) Distance Reciprocal Distortion (DRD)*:

$$DRD = \frac{\sum_k DRD_k}{NUBN} \tag{4}$$

where $NUBN$ and DRD_k are the number of nonuniform (not entirely black or white pixels) 8×8 patches in the GT image and the distortion of the k-th flipped pixel, respectively. This metric evaluates the extent of visual distortion in binarized document images.

III. BACKGROUND

In this section, we provide a brief description of the wavelet and three CNN architectures (SegNet, U-net and DeepLabv3+) concepts.

A. WAVELETS

In our implementation, we use the wavelet transform in terms of high-pass and low-pass functions. Compared with other transforms, structures such as edges, endpoints and vertices are better preserved and can be used as a noise removal tool. Additionally, edges are well isolated in different orientations and subbands, which leads to superior results in image processing applications. To implement the 2D wavelet transform in separable rows and columns, we use a scaling or approximation function $\psi(x,y)_{cA}$ and three wavelet or detail functions $\psi(x,y)_{cH}$, $\psi(x,y)_{cV}$, $\psi(x,y)_{cD}$ based on [8], [53]:

$$\psi(x,y)_{cA} = \phi(x)\phi(y) \quad (LLwavelet), \tag{5}$$
$$\psi(x,y)_{cH} = \phi(x)\psi(y) \quad (LHwavelet), \tag{6}$$
$$\psi(x,y)_{cV} = \psi(x)\phi(y) \quad (HLwavelet), \tag{7}$$
$$\psi(x,y)_{cD} = \psi(x)\psi(y) \quad (HHwavelet) \tag{8}$$

The LL, LH, HL and HH wavelets (L and H correspond to low frequency and high frequency) are defined as the product of the low-pass $\phi(.)$ and the high-pass $\psi(.)$ functions along the first and the second dimensions, respectively. It should be noted that the LL wavelet is an approximation coefficient, and the LH, HL and HH wavelets are the horizontal, vertical and diagonal detail coefficients, respectively. To implement the wavelet discretely, that is, the discrete wavelet transform (DWT) for image $I(x,y)$ is:

$$W_{\phi}^{cA}(j_0,a,b) = \frac{1}{\sqrt{AB}} \sum_{x=0}^{A-1} \sum_{y=0}^{B-1} I(x,y) \phi_{j_0,a,b}^{cA}(x,y), \tag{9}$$

$$\boldsymbol{W}^{t}_{\psi}(j,a,b) = \frac{1}{\sqrt{AB}} \sum_{x=0}^{A-1} \sum_{y=0}^{B-1} I(x,y)\psi^{t}_{j,a,b}(x,y) \quad (10)$$

where

$$\phi_{j_0,a,b}(x,y) = 2^{j_0/2}\phi(2^{j_0}x - a, 2^{j_0}y - b) \quad (11)$$

$$\psi^{t}_{j,a,b}(x,y) = 2^{j/2}\psi^{t}(2^{j}x - a, 2^{j}y - b) \quad (12)$$

t is set of horizontal, vertical and diagonal directions $t = \{cH, cV, cD, \}$. j_0 tunes the scale of the scaling function initially. Given $j \geq j_0$, $\boldsymbol{W}^{cA}_{\phi}(j_0,a,b)$ and $\boldsymbol{W}^{t}_{\psi}(j,a,b)$ show the coefficients of the approximation and the details of the horizontal, vertical, and diagonal for $I(x,y)$, respectively. $j_0 = 0$, $A = B = 2^J$, $j = 0, 1, 2, \ldots, J-1$ and $a, b = 0, 1, 2, \ldots, 2^j - 1$ are given. When given $\boldsymbol{W}_{\phi}$ and $\boldsymbol{W}^{t}_{\psi}$, we can approximate $I(x,y)$ via the inverse discrete wavelet transform:

$$I(x,y) = \frac{1}{\sqrt{AB}} \sum_{a} \sum_{b} \boldsymbol{W}_{\phi}(j_0,a,b)\phi_{j_0,a,b} + \frac{1}{\sqrt{AB}} \sum_{t=cH,cV,cD} \sum_{j=j_0}^{\infty} \sum_{a} \sum_{b} \boldsymbol{W}^{t}_{\psi}(j,a,b)\psi^{t}_{j,a,b} \quad (13)$$

B. CNNs ARCHITECTURES

Badrinarayanan *et al.* [54] introduced SegNet as a novel architecture of the convolutional neural network (CNN) for semantic image segmentation. Typically, the predominant structure of SegNet is constituted of an encoder network, decoder network (corresponding to each encoder) and a final pixelwise classification layer. The architecture structure is shown in Fig. 3. Given a depth of 4 for the network, 13 convolutional layers are considered in the encoder network, corresponding to the number of convolutional layers in the VGG16 network [55]. For multiclass classification, a softmax classifier connected to the final decoder predicts each pixel of the image independently. According to the pooling process on the encoder network side, input feature map(s) are upsampled from the equivalent encoder feature map(s) on the other side, and these processes lead to a sparse feature map(s).

A U-net [49] is an architecture with a contradicting path (downpath) to acquire context and a symmetrically extending path (uppath) to enable granular localization. The contradicting path starts with the convolutional layers that follow with pooling actions and iteratively downsamples feature maps. Each step in the extensive path starts with an upsampling of the feature map proceeded by a convolution. It is worthy to note that compared with U-net, SegNet reuses the pooling indices and does not transfer the entire feature map to reduce the memory cost.

To study the impact of our approach on newer architectures, we consider DeepLabv3+ as a powerful deep network that was recently developed. DeeplabV3+ [56], as an extended DeepLabv3 network [57], was introduced to refine the object boundaries by adding a decoder. The encoder-decoder network utilizes the convolutional neural network, called Xception with atrous convolution layers, to acquire encoder features. The combination of the Xception model and atrous separable convolution makes the model faster and stronger.

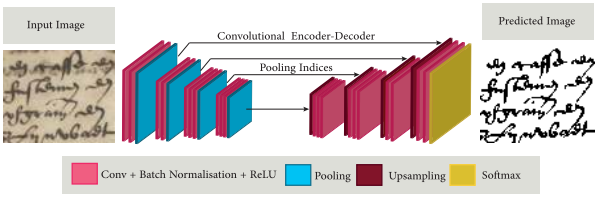

FIGURE 3. Overview of the SegNet architecture.

IV. CNNs AND WAVELET-BASED MULTICHANNEL IMAGES

This section proposes our approach for degraded document binarization. First, the databases used to create the training and testing sets are introduced to correctly classify text pixels in the degraded document images. Our approach is primarily based on SegNet, U-net and DeepLabv3+, and wavelet-based multichannel images are described. Finally, to show our motivation to choose the wavelets, we visualize our features during training.

A. DATABASE

The public database introduced in [5][1] is used to create our multichannel images for the training set. Patches of the database are generated by several databases, i.e., ICDAR 2009 [4], ICFHR 2010 [40], ICFHR 2012 [42], the Persian heritage image binarization database (PHIDB) [58], the Bickley-diary database [59] and the Synchromedia multispectral database (S-MS) [60]. For testing sets, we used the ICDAR 2011 [41] (DIBCO 2011 database), ICDAR 2013 [43] (DIBCO 2013 database), ICFHR 2014 [44] (H-DIBCO 2014 database), ICFHR 2016 [46] (H-DIBCO 2016 database), ICDAR 2017 [48] (DIBCO 2017 database), ICFHR 2018 [50] (H-DIBCO 2018 database), and ICDAR 2019 [51] (DIBCO 2019) competition databases. Data augmentation is performed on the extracted input image by applying rotation with angles of 90, 180, or 270 to produce 84210 patches of the training image. In the database, to create image patches, a local window of size $a \times a$ (that is, the proportion of the width to the height in each image, $R = width/height$) was selected to touch all image regions. The size a is obtained as follows:

$$a = \begin{cases} min(width, height)/2 & if\ 0.5 < R < 2 \\ max(width, height) & if\ R \leq 0.5\ or\ R \geq 2 \end{cases} \quad (14)$$

Then, a distance was selected with overlap $a/2$ in both ways. In our approach, all patches were finally normalized to a 192×192 pixel size due to the memory drawback. Fig. 4 shows samples of created patches and their ground truth.

[1] Available at https://github.com/vqnhat/ DSN-Binarization/

FIGURE 4. Samples of created patches and their ground truth.

In theory, instead of image patches, the whole image can be fed into the CNNs. However, we had the limitation of the GPU in the experimental environment with arbitrary inputs. To sort the network model with a large database, additional GPU memory was required, and when the size of the input images increased, the number of pooling layers was accordingly increased. Additionally, when used in CNNs for segmentation, image patches can provide enhanced accuracy and labeling consistency by capturing contextual information [5].

B. WAVELET-BASED MULTICHANNEL IMAGES

In this section, the implementation of multichannel images based on the wavelet is presented. Fig. 5 shows samples of the images based on our approach. It should be noted that different features obtained at different frequencies (especially high-frequency details) play an important role in image representation [61]. CNNs suffer from the drawback of simultaneously retaining high- and low-frequency detail. This drawback leads us to use the advantages of multispectral images, that is, supplying additional information on each pixel for semantic segmentation purposes with promising results [38]. Additionally, in document binarization, wavelet processing can be used as an enhancement step to preserve the foreground strokes [62], and transformation and kernels are also utilized to extract other features from the original images; those extracted features are used to reconstruct the text [63]. The obtained features can be used in the convolution layers for preserving edges [64]. These factors finally inspired us to create multichannel images in different frequency bands. To produce the images, initially, the wavelet transform upon the original image $I(x, y)$ decomposes the image into one subband with low frequency (cA) and three subbands with high frequency (cH, cV and cD), as shown in Fig. 5(a). The approximation process is based on setting three of the four subbands to zero and subsequently computing the single-level reconstructed approximation coefficients based on the approximation coefficients of one of the four subbands. The process is formulated by using equation (13):

$$I_{Ap}^{cA}(x,y) = \frac{1}{\sqrt{AB}} \sum_{a} \sum_{b} \boldsymbol{W}_{\phi}(j_0,a,b)\phi_{j_0,a,b}^{cA}(x,y) \tag{15}$$

$$I_{Ap}^{cH}(x,y) = \frac{1}{\sqrt{AB}} \sum_{t=cH} \sum_{j=j_0}^{\infty} \sum_{a} \sum_{b} \boldsymbol{W}_{\psi}^{t}(j,a,b)\psi_{j,a,b}^{cH}(x,y) \tag{16}$$

$$I_{Ap}^{cV}(x,y) = \frac{1}{\sqrt{AB}} \sum_{t=cV} \sum_{j=j_0}^{\infty} \sum_{a} \sum_{b} \boldsymbol{W}_{\psi}^{cV}(j,a,b)\psi_{j,a,b}^{t}(x,y) \tag{17}$$

$$I_{Ap}^{cD}(x,y) = \frac{1}{\sqrt{AB}} \sum_{t=cD} \sum_{j=j_0}^{\infty} \sum_{a} \sum_{b} \boldsymbol{W}_{\psi}^{cD}(j,a,b)\psi_{j,a,b}^{t}(x,y) \tag{18}$$

where the images approximated by coefficients cA, cH, cV and cD are I_{Ap}^{cA}, I_{Ap}^{cH}, I_{Ap}^{cV} and I_{Ap}^{cD}, respectively. Ap shows that these images are approximated by the coefficients (inverse discrete wavelet transform). It should be noted that we use the original image $I(x, y)$ instead of the image approximated by coefficients cA (I_{Ap}^{cA}) because both the images have the same data (low frequency). Therefore, the original image is assembled with the image approximation based on the coefficients of three subbands (high frequencies).

Finally, the images are prepared for two approaches: single- and multiple-stream networks. For the single-stream network, the input image with four channels is created, namely, the original image and the three subbands that approximate the original image (Fig. 5(b)). For the latter approach, we consider three images with two channels as the input of multiple networks (Fig. 5(c)). The approach is applied to investigate the impact of each subband in the image segmentation (document binarization). It should be noted that the information obtained from the wavelet transform is not the same original image, and it contains additional pieces of information that are added to the original channel. The information is the approximation of the original image together with horizontal, vertical and diagonal details.

The method provided for generating wavelet-based multichannel images is summarized as detailed below:

- Decomposition of the image (given a grayscale document image) into separate cA, cH, cV and cD subbands.
- Approximation of the original image with each separate subband (I_{Ap}^{cA}, I_{Ap}^{cH}, I_{Ap}^{cV} and I_{Ap}^{cD}).
- Arrangement and development of a multichannel image of 4 channels for single networks with regards to the four-channel width-by-height arrays (in which the width and height of the original image are the width and height, respectively). The original grayscale image ($I(x, y)$) is

FIGURE 5. (a) Four wavelet subbands of the document image of the DIBCO 2017 database. (b) and (c) are samples of the wavelet-based multichannel images (displayed wavelet channels are based on scaled colors).

the first image channel (instead of I_{Ap}^{cA}), and the image approximated by the three high frequency subbands is the 2nd (I_{Ap}^{cH}), 3rd (I_{Ap}^{cV}) and 4th (I_{Ap}^{cD}) image channels.

- Arrangement and creation of three multichannel images with 2 channels for multiple networks of width-by-height two-channel arrays. The first output image

includes the original grayscale image ($I(x, y)$), which is the 1st image channel, and the image approximated I_{Ap}^{cH} by the subband cH is the 2nd image channel. The arrangement, that is, $I(x, y)$ *and* I_{Ap}^{cH} is called C_1. The same conditions are repeated for the second and third images ($I(x, y)$ and I_{Ap}^{cV}, and $I(x, y)$ and I_{Ap}^{cD} are called C_2 and C_3, respectively).

C. SINGLE AND MULTIPLE NETWORKS FOR DOCUMENT BINARIZATION

The segmentation process with single-stream CNNs on the original images when lower-level data such as character contours and edges are processed leads to incomplete results in the binarization problem [5]. Additionally, in this research, our methods are based on increasing performance with respect to the state-of-the-art methods.

To improve on the previous attempts, we propose two architectures that can obtain the details of the images by simultaneously considering their high frequencies and original images. We use two approaches based on CNNs to train the multichannel images based upon the wavelet (single and multiple networks). We select SegNets, U-net and DeepLabv3+ (as three of the accomplished CNNs for semantic segmentation). These two approaches for the SegNet architecture are shown in Fig. 6. In the first approach, we use the images created in the previous section with four channels for input into the single network. In this approach, we simultaneously consider all images approximated with subbands with high frequencies in accordance with the original image. This approach is selected to incorporate both low-level features (such as edges or contrasts) and high-level features (such as the object area) and extract them in a single-stream network.

In the second approach, we consider three networks to construct each individual network with the input image based on two channels (original image and one of the images approximated with the subbands with high frequency). The predicted maps at each level are based on the multiples, and each stream of the network has the ability to separately discriminate special features and criteria on different levels. To create the final binary image, prediction of every pixel as text or background is performed and integrated from the trained networks at different feature streams. Given P_1, P_2, and P_3 as the three predicted maps for each image patch, the minimum function (Min) is applied to integrate the information. This function can be used as a noise pixel removal tool with a high-probability value. The final predicted map is based on the following [5]:

$$P\_F = min(P\_1, P\_2, P\_3). \tag{19}$$

It should be noted that images in the testing set are processed with the same patch size as that of the training set. If the considered patches overlap, then we select the maximum value based on [5] to solve the weak response problem at the boundary regions.

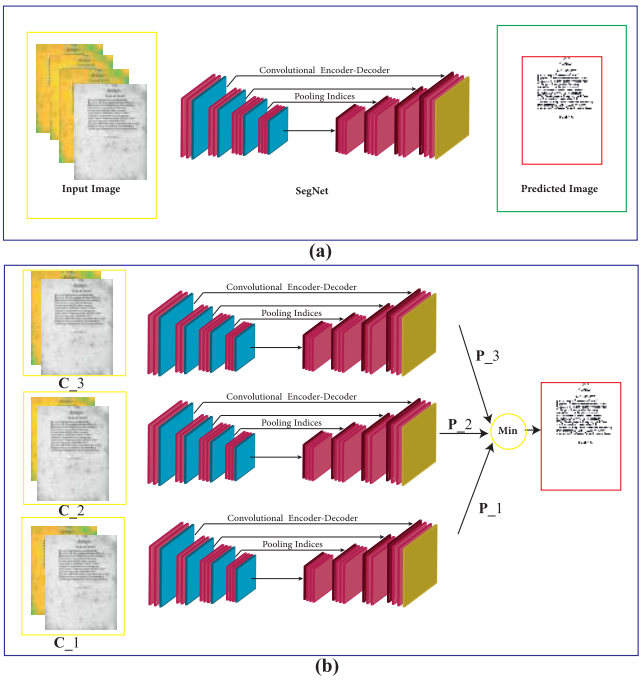

FIGURE 6. Single and multiple architectures based on wavelet-based multichannel images.

D. FEATURE VISUALIZATION

To show the impact of our created multichannel images based on wavelets that are fed into the CNNs to learn the discriminative features and classify one class from another, we visual the process of creating the features based on original and multichannel images during the feature extraction process and after a convolutional layer. The utility of using the wavelet low- and high-frequency subbands is to eliminate the noise and maintain the feature map structures simultaneously. The noise in the image will impact features generated by CNNs [65]. At the same time, we would like to retain the function map structures. For the segmentation function, these structures are quite relevant.

To show the features after a convolutional layer, we consider a patch with a size of 48 × 48 for original and multichannel images as input. To produce feature maps in the convolutional layer, let 6 convolutional kernels with sizes of 5 × 5 obtain 6 feature maps with sizes of 44 × 44, as illustrated in Fig. 7 (a). To better understand CNNs based on multichannel images, we have shown original and multichannel image patches in Fig. 7 (b) with obvious structures. Fig. 6 (c) visualizes the results of the original channel and channels created by the wavelet transform. The results are the output of the first convolution layer of CNN when the input is one of the channels (original image in grayscale mode and channels approximated with wavelet subbands). The networks are fed by each channel in the experiment separately.

It is observed that CNN based on multichannel images can learn the additional features compared with only original images. It is worth noting that, based on the type of the input images, the features learned by our approach retain more

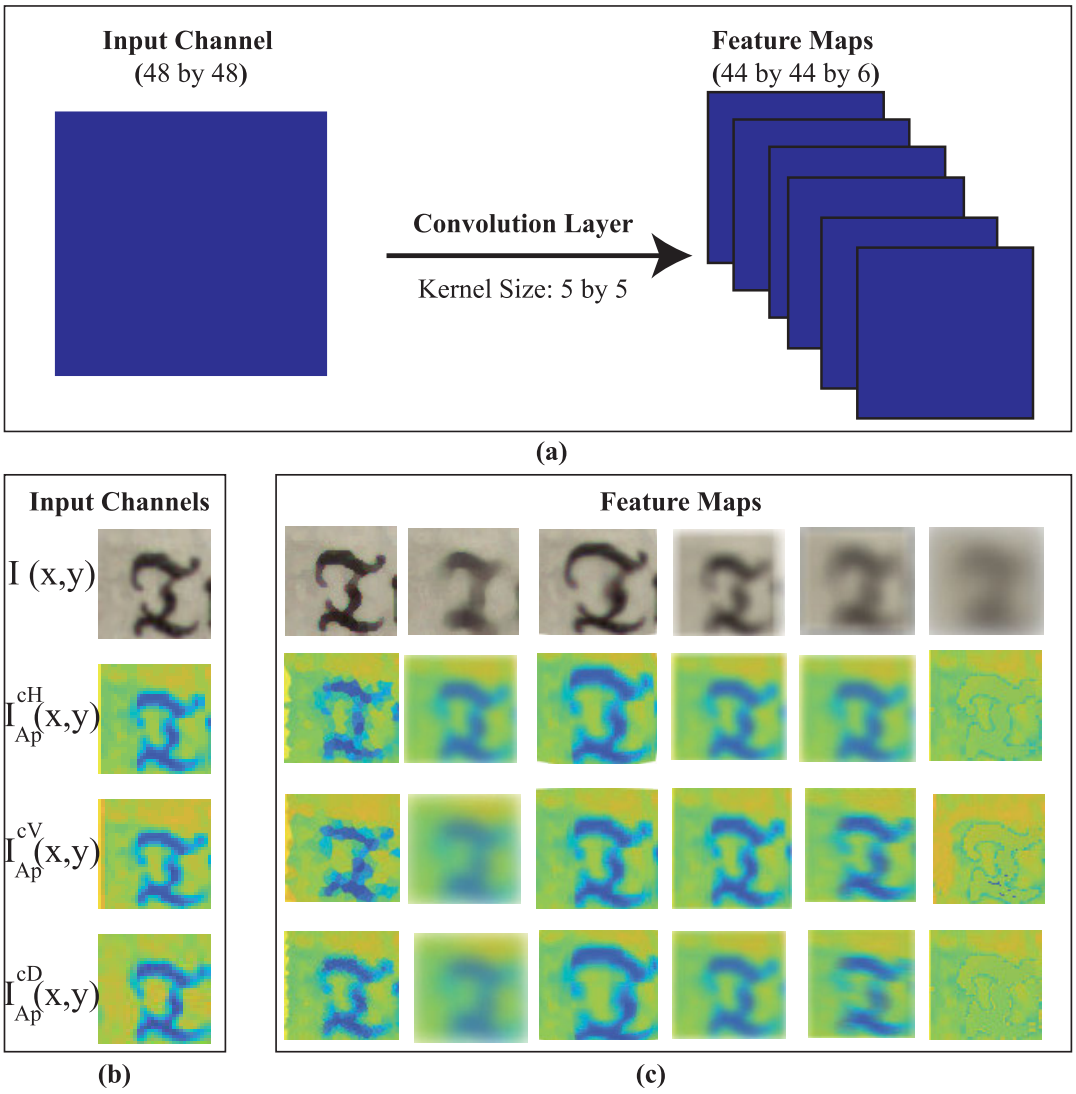

FIGURE 7. Process of extracting features: (a) Convolutional layer outputs in a network architecture. (b) Feature maps obtained from the $I(x, y)$, I_{Ap}^{cH}, I_{Ap}^{cV} and I_{Ap}^{cD} channels.

structures than those of CNNs. Additionally, the features learned by our approach have a better visualization in comparison to the features learned by CNN based on original images. The significance is that the wavelet increases the efficiency of the network. The explanation is that the wavelet not only keeps the structures of the extracted features of historical documents but it also eliminates any noise during the learning process.

V. EVALUATION

The experimental results from our proposed approaches are presented in this section. Additionally, we explain the execution of the document image binarization methods. State-of-the-art binarization methods are considered for comparison purposes. Finally, we provide more analysis of our approaches. Implementation details and source code with the selected databases and trained models are freely available at: https://github.com/YounesAkbari/Document-Binarization.

A. IMPLEMENTATION AND SETTING

We train two approaches independently over the created wavelet-based multichannel image patches. The images were decomposed into a series of subbands using the Daubechies-2 (db2) wavelet due to the wide variety of singularities of the interfering strokes [62]. To set the wavelet decomposition levels, experiments showed that a one-level decomposition is sufficient and that the use of multiple levels does not have any impact on the segmentation rates. The U-net and SegNet layers are set with an encoder/decoder depth of 3 and DeepLabV3+ with a default setting. As stated previously, our networks contain an encoder and a related decoder subnetwork. Selecting the depth of the encoder/decoder networks during the learning process determines the number of instances where the input image is downsampled or upsampled. Additionally, 50 mini-batches were trained at each step; the maxEpochs was set to 10. We enacted all simulations in MATLAB R2019a. To run all the experiments, we used a

FIGURE 8. Binarization outputs for HW1 of the DIBCO 2011 database: (a) original image, (b) ground truth, (c) Otsu [13], (d) DeepLabv3+ (with original images), (e) our method (single network), (f) our method (multiple networks), (g) U-net (with original images), (h) our method (single network), (i) our method (multiple networks), (j) SegNet (with original images), (k) our method (single network), and (l) our method (multiple networks).

FIGURE 9. Binarization outputs for HW7 of the DIBCO 2013 database: (a) original image, (b) ground truth, (c) Otsu [13], (d) DeepLabv3+ (with original images), (e) our method (single network), (f) our method (multiple networks), (g) U-net (with original images), (h) our method (single network), (i) our method (multiple networks), (j) SegNet (with original images), (k) our method (single network), and (l) our method (multiple networks).

64-bit operating system with a CPU E5-2690 v3 @ 2.60 GHz, 64.0 GB of RAM and a single NVIDIA GTX TITAN X. Furthermore, to explore the impact of the multichannel images, the original training images (one channel) were additionally trained on the SegNet, U-net and DeepLabv3+ networks.

B. COMPARISON WITH OTHER BINARIZATION METHODS

Seven databases are selected to assess the proposed approaches, namely, ICDAR 2011 (DIBCO 2011 database), ICDAR 2013 (DIBCO 2013 database), ICFHR 2014 (H-DIBCO 2014 database), ICFHR 2016 (H-DIBCO 2016 database), the ICDAR 2017 (DIBCO 2017 database), ICFHR 2018 (H-DIBCO 2018 database), and ICDAR 2019 (DIBCO 2019 database) competition databases. Note that three databases, H-DIBCO 2014, H-DIBCO 2016 and H-DIBCO 2018, only contain handwritten images, whereas the other four databases, the DIBCO 2011, DIBCO 2013, DIBCO 2017, and DIBCO 2019 contain both machine-printed and handwritten images. The state-of-the-art methods that are compared with our proposed approaches are the traditional methods [15], [24], [27], [66]–[68], deep learning-based methods [5], [10], [37] and the winner method in each of the seven competitions. It should be noted that some deep learning methods, such as [34], were evaluated solely on one metric and that generally, deep learning methods are the difference in training data. Therefore, a fair comparison with the methods is considerably difficult to execute. For all of the methods with publicly available code used on the testing databases, the default parameters mentioned in their codes were retained. To assess the methods, we utilized four evaluation metrics: FM, F_{ps}, PSNR and DRD. It should be noted that we evaluated our experiments based on the metric files (DIBCO_metrics.exe) provided in each competition by its organizers. Subsection II.C includes a detailed description of the measures. Furthermore, our quality of binarization results is shown in Figs. 8 to 14. The quantitative results, in terms of FM, F_{ps}, PSNR and DRD measures, from all databases are shown in Tables 2 to 5. According to the evaluations, the three networks with multichannel images yield better quantitative and qualitative results in binarization compared with the original images. Additionally, in terms of the three databases, DIBCO 2013, H-DIBCO 2014 and DIBCO 2017, our results reveal that the two approaches outperform the methods.

1) QUANTITATIVE EVALUATION

The quantitative results in terms of FM, F_{ps}, PSNR and DRD measured from all databases are shown in Tables 2 to 5. The tables reveal that our technique executes considerably better than all of the classic binarization techniques on the DIBCO 2011 database and is as sound as the DSN method (based on CNNs) [5] with respect to all measures.

The quantitative results from DIBCO 2013 show that the proposed method (single and multiple networks) obtains the

TABLE 2. Comparison of our approach based on multichannel single networks (MSN) and multichannel multiple networks (MMN) with other binarization methods in terms of FM measure on DIBCO and H-DIBCO (2011 to 2019) image sets (highest scoring values highlighted in bold).

Year	Winner	[14]	[66]	[30]	[37]	[67]	[23]	[68]	[26]	[5]	U-net	Our MSN	Our MMN	SegNet	Our MSN	Our MMN	Deep Labv3+	Our MSN	Our MMN
2011	88.70	71.30	84.30	**93.60**	91.87	87.80	91.70	92.48	91.92	93.30	90.25	91.96	92.62	91.68	92.16	93.06	92.86	93.40	93.55
2013	92.10	85.00	83.40	93.17	93.97	87.70	91.30	90.78	93.42	94.40	94.25	95.22	**95.47**	94.30	95.15	95.43	94.00	94.70	95.15
2014	96.88	86.83	91.48	91.96	93.79	94.38	96.49	96.14	94.98	96.66	94.40	95.76	96.27	94.83	96.10	96.95	95.30	95.70	**97.05**
2016	87.61	82.52	-	89.52	90.18	84.75	87.47	87.21	**90.48**	90.10	85.60	87.95	88.94	85.58	88.10	89.21	88.25	89.90	90.15
2017	**91.04**	77.11	-	-	-	87.80	91.70	-	-	-	87.15	88.27	88.58	87.00	88.20	88.34	88.65	90.18	90.85
2018	88.34	67.81	-	-	-	-	-	-	-	-	79.85	80.26	81.38	79.18	82.94	84.90	84.85	87.40	**89.05**
2019	**72.85**	42.51	-	-	-	-	-	-	-	-	56.37	57.76	58.64	61.95	62.68	63.75	61.97	62.70	65.54

TABLE 3. Comparison of our approach based on multichannel single networks (MSN) and multichannel multiple networks (MMN) with other binarization methods in terms of F_{ps} measure on DIBCO and H-DIBCO (2011 to 2019) image sets (highest scoring values highlighted in bold).

Year	Winner	[14]	[66]	[30]	[37]	[67]	[23]	[68]	[26]	[5]	U-net	Our MSN	Our MMN	SegNet	Our MSN	Our MMN	Deep Labv3+	Our MSN	Our MMN
2011	-	71.30	84.30	**97.70**	-	90.00	92.00	94.11	95.09	96.40	92.46	94.00	95.15	94.80	95.19	96.01	95.63	96.25	96.45
2013	94.20	89.80	87.00	96.81	-	88.30	91.70	91.47	96.05	96.00	95.95	96.79	96.91	95.83	96.88	**97.02**	95.55	96.80	96.67
2014	97.65	91.80	95.48	94.78	-	95.94	97.38	96.73	97.18	97.59	95.00	96.15	97.05	95.45	96.90	**97.92**	95.90	96.85	97.55
2016	91.28	86.85	-	93.76	-	88.94	92.28	88.48	93.27	93.57	89.63	91.20	93.21	90.61	92.27	**93.63**	91.00	91.75	93.46
2017	92.86	84.10	-	-	-	90.00	92.00	-	-	-	87.15	92.20	93.50	89.87	91.12	93.45	91.20	93.55	**93.60**
2018	90.24	74.08	-	-	-	-	-	-	-	-	87.97	91.75	93.54	84.26	88.41	91.81	89.25	92.64	**93.65**
2019	**72.15**	39.76	-	-	-	-	-	-	-	-	54.43	55.95	56.81	57.52	60.28	61.72	61.62	62.39	64.19

TABLE 4. Comparison of our approach based on multichannel single networks (MSN) and multichannel multiple networks (MMN) with other binarization methods in terms of PSNR measure on DIBCO and H-DIBCO (2011 to 2019) image sets (highest scoring values highlighted in bold).

Year	Winner	[14]	[66]	[30]	[37]	[67]	[23]	[68]	[26]	[5]	U-net	Our MSN	Our MMN	SegNet	Our MSN	Our MMN	Deep Labv3+	Our MSN	Our MMN
2011	17.80	12.50	16.30	**20.11**	19.07	17.70	19.30	19.37	18.98	20.10	17.75	18.12	19.13	19.01	19.32	19.73	18.19	19.97	20.10
2013	20.70	16.90	17.10	20.71	21.30	19.60	21.30	20.54	20.78	21.40	19.81	22.17	22.25	20.69	22.15	**22.33**	20.90	22.00	22.15
2014	22.66	17.63	18.54	20.76	20.79	20.31	22.24	22.24	20.56	**23.23**	20.75	21.46	21.97	20.98	22.05	22.52	21.40	22.00	21.85
2016	18.11	16.42	-	18.67	18.99	17.64	18.05	17.36	**19.30**	19.01	17.00	17.88	18.33	16.95	18.02	18.39	18.55	19.18	19.25
2017	18.28	14.25	-	-	-	17.70	**19.30**	-	-	-	16.85	17.32	17.85	16.71	17.18	17.04	17.20	18.24	18.50
2018	19.11	13.78	-	-	-	-	-	-	-	-	16.00	16.25	16.20	17.33	16.62	16.86	18.40	19.10	**19.17**
2019	**14.47**	7.71	-	-	-	-	-	-	-	-	12.61	12.87	13.09	11.50	12.33	12.70	11.89	13.38	12.95

TABLE 5. Comparison of our approach based on multichannel single networks (MSN) and multichannel multiple networks (MMN) with other binarization methods in terms of DRD measure on DIBCO and H-DIBCO (2011 to 2019) image sets (highest scoring values highlighted in bold).

Year	Winner	[14]	[66]	[30]	[37]	[67]	[23]	[68]	[26]	[5]	U-net	Our MSN	Our MMN	SegNet	Our MSN	Our MMN	Deep Labv3+	Our MSN	Our MMN
2011	8.67	24.10	6.08	**1.85**	2.57	4.70	3.48	2.97	2.64	2.00	3.79	3.32	3.00	3.06	2.70	2.21	2.80	2.18	1.95
2013	3.10	7.60	9.50	2.21	1.83	4.20	3.20	3.59	2.03	1.80	2.86	2.10	1.76	2.31	1.54	**1.46**	2.60	2.13	2.05
2014	0.90	4.89	2.73	2.72	2.30	1.95	1.08	1.25	1.50	**0.79**	2.64	2.13	1.21	2.08	1.50	0.90	2.30	1.32	1.02
2016	3.86	7.49	-	3.76	3.61	5.64	5.35	5.27	3.97	**3.58**	5.10	4.60	3.87	5.12	4.45	3.68	4.03	3.80	3.63
2017	3.40	8.85	-	-	-	4.70	3.48	-	-	-	4.46	4.05	3.42	4.99	4.52	**3.27**	3.90	3.45	3.30
2018	4.92	17.69	-	-	-	-	-	-	-	-	6.80	6.20	5.74	6.99	5.34	5.59	5.20	4.85	**4.80**
2019	**16.23**	112.40	-	-	-	-	-	-	-	-	18.07	17.11	16.63	21.30	20.45	19.57	18.82	17.71	17.26

first rank in terms of all the measurements. As noted from the low value of the DRD, the visual distortion is also more powerful than the others in the proposed method. Compared with the deep learning-based methods (FCN [10], PDNet [37], and DSN [5]), our approach obtains the highest score.

Our method outperforms all of the traditional binarization methods on H-DIBCO 2014 in terms of the four metrics. Our method is also more effective with respect to the two measures (FM and F_{ps}) compared with all the other methods. In the other two evaluation metrics, our approach comes in second place after the DSN method [5].

Compared with the other methods, our quantitative results from H-DIBCO 2016 obtain third-place with regards to the three evaluation measures. The DSN and SSP methods both generate better results than our approach in the three measures of FM, PSNR and DRD. The explanation for this outcome is probably that the stroke width of the handwritten image is no longer uniform, which results in a faulty estimation of the stroke width. As the F_{ps} metric shows, our system performs better in the preservation of damaged and incomplete texts compared to other methods.

As shown in the tables, our method achieves first place in terms of the F_{ps} and DRD measures on DIBCO 2017 database. For the two remaining measures, the winner of the competition and the Howe method are better than our method.

As shown in the report of the competition H-DIBCO 2018 [50], a traditional method (based on Howe method) obtained the best rank, and as shown in the tables, the database presented a challenge for the deep learning traditional techniques such as SegNet and U-net that cannot outperform the winner. As shown in the table, our method based on DeepLabv3+ obtains the first place in terms of the four measures. However, SegNet and U-net, based on multichannel images, fully outperform the two networks when using the original image.

The last competition for both handwritten and machine-printed documents is DIBCO 2019 [50], in which the winning method does not use a deep learning approach (based on three clustering algorithms, namely Fuzzy C-Means, K-Medoids, and K-Means++). The results reported in the competition show that the database is indeed very challenging

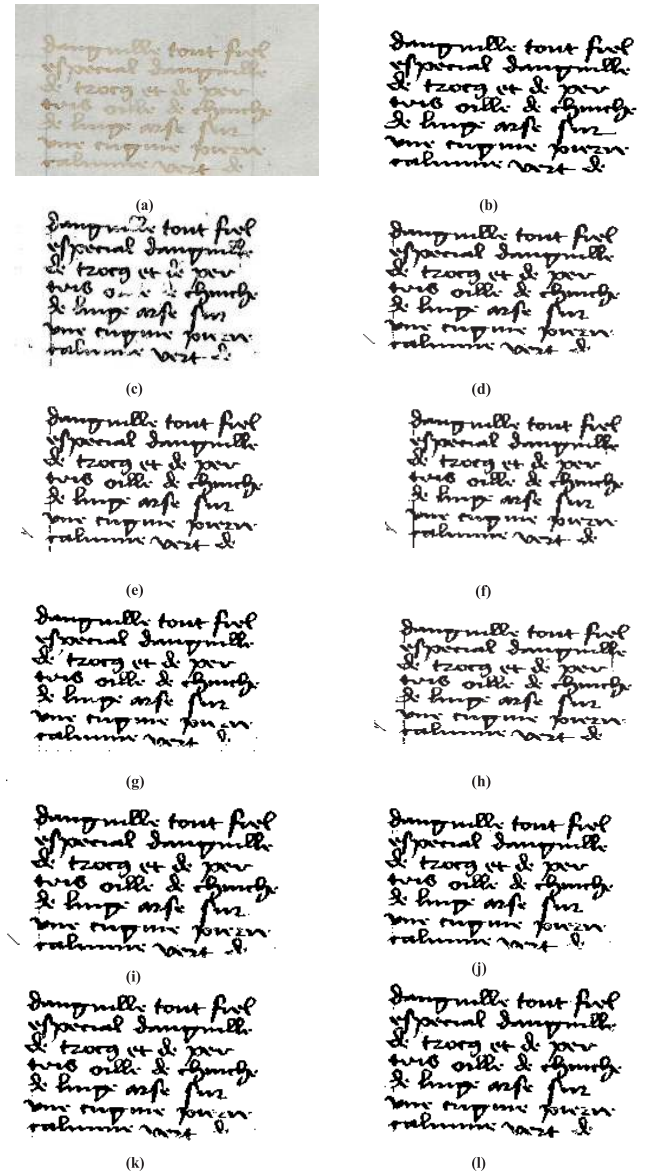

FIGURE 10. Binarization outputs for H06 of the DIBCO 2014 database: (a) original image, (b) ground truth, (c) Otsu [13], (d) DeepLabv3+ (with original images), (e) our method (single network), (f) our method (multiple networks), (g) U-net (with original images), (h) our method (single network), (i) our method (multiple networks), (j) SegNet (with original images), (k) our method (single network), and (l) our method (multiple networks).

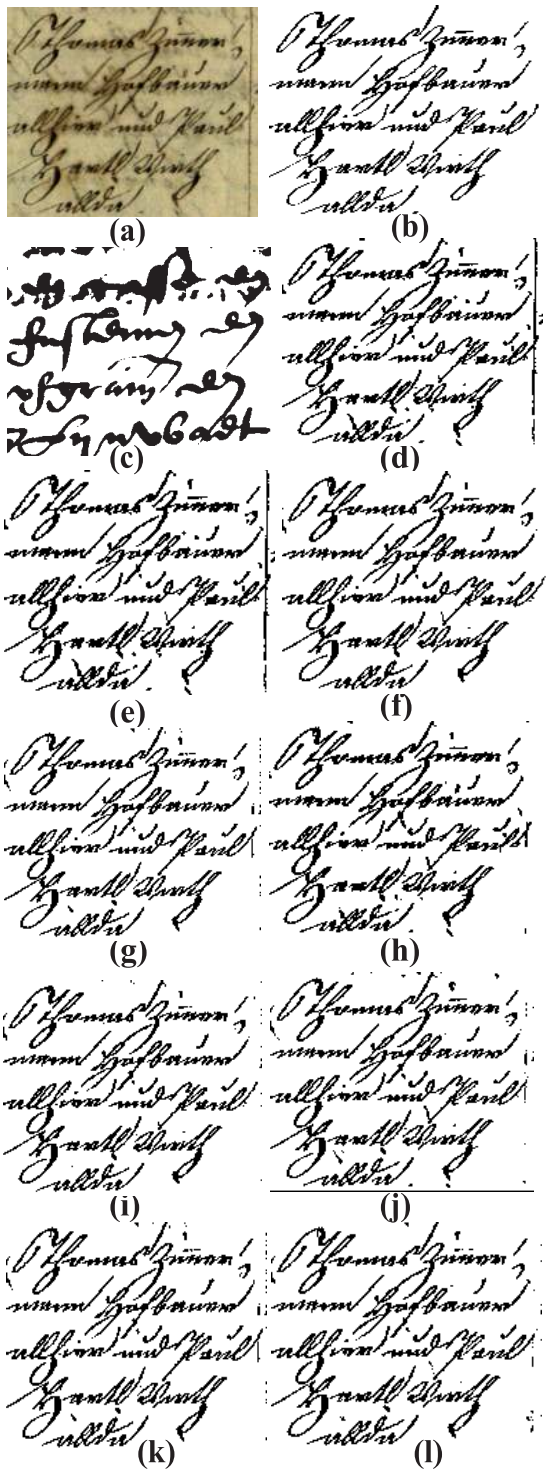

FIGURE 11. Binarization outputs for the challenging image (10) of the DIBCO 2016 database: (a) original image, (b) ground truth, (c) Otsu [13], (d) DeepLabv3+ (with original images), (e) our method (single network), (f) our method (multiple networks), (g) U-net (with original images), (h) our method (single network), (i) our method (multiple networks), (j) SegNet (with original images), (k) our method (single network), and (l) our method (multiple networks).

and therefore improvement is necessary. As shown in the tables, the results obtained by the winner are better than our approach, and this proves that deep learning methods are very sensitive to training databases. However, all three architectures (DeepLabv3+, SegNet, and U-net), based on multichannel images, fully outperform the three architectures when using the original image.

2) QUALITATIVE EVALUATION

Certain binary results processed on an example image (HW1) of the DIBCO 2011 image set are shown in Fig. 8. As shown by the results, the three methods based on the original images achieve clean binary images. However, the result of the proposed method based on multichannel images in both

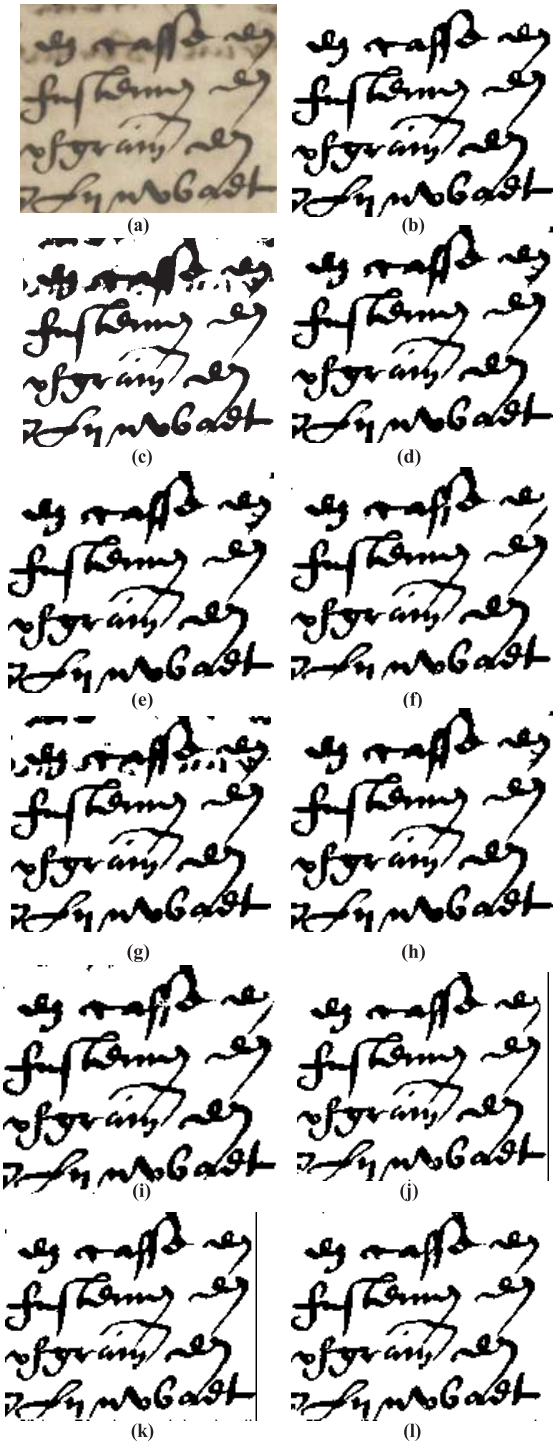

FIGURE 12. Binarization outputs for the challenging image (6) of the DIBCO 2017 database: (a) original image, (b) ground truth, (c) Otsu [13], (d) DeepLabv3+ (with original images), (e) our method (single network), (f) our method (multiple networks), (g) U-net (with original images), (h) our method (single network), (i) our method (multiple networks), (j) SegNet (with original images), (k) our method (single network), and (l) our method (multiple networks).

single and multiple networks is more effective. Our proposed method, with simultaneous use of the high and low frequencies in the input image, can efficiently produce a higher visual quality from the input image with a dark shadow. As explored

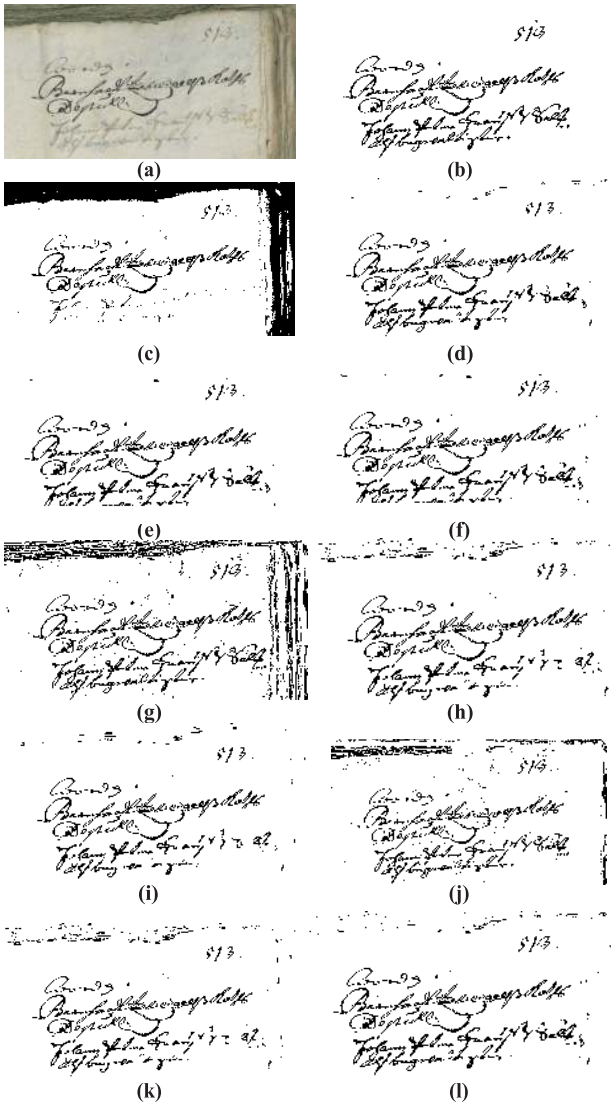

FIGURE 13. Binarization outputs for the challenging image (5) of the H-DIBCO 2018 database: (a) original image, (b) ground truth, (c) Otsu [13], (d) DeepLabv3+ (with original images), (e) our method (single network), (f) our method (multiple networks), (g) U-net (with original images), (h) our method (single network), (i) our method (multiple networks), (j) SegNet (with original images), (k) our method (single network), and (l) our method (multiple networks).

within the literature, e.g., [5], arduous cases (weaknesses mentioned in the introduction section) are included in the DIBCO 2013 database. The effects of demanding text-like backgrounds are illustrated in Fig. 9 (HW07). As exhibited, the three networks with original images are successful in separating the text, which serves as the foreground, from the background, but they also fail in preserving thin strokes. The proposed approach accomplishes the best visual quality for the sample.

The H-DIBCO 2014 database consists only of historical handwritten documents. In the qualitative assessment, as shown in Fig. 10, our method is smooth in keeping the text in the sample. The H-DIBCO 2016 database (sample shown in Fig. 11) is also a challenging database in that

additional improvements are required compared with the results of H-DIBCO 2014. The DeepLabv3+ network with multichannel images generates better results than other networks. A total of 20 images (10 handwritten and 10 machine-printed images) are used to test the algorithms in the DIBCO 2017 database. Fig. 12 presents the binarization results produced by the three networks (with original and without multichannel images). As shown, our approach is also better than the networks with the original images with regard to visual distortion and broken and missing text. Additionally, our method based on DeepLabv3+ outperforms other methods in terms of visual distortion and broken and missing text. The H-DIBCO 2018 database (sample is shown in Fig. 13) has 10 handwritten images. As shown in the figure, the DeepLabv3+ method based on multichannel images obtains a clean background compared to other approaches. The DIBCO 2019 database includes 20 images (10 handwritten and 10 machine-printed images.) Fig. 14 shows that the DeepLabv3+ method based on multichannel images obtains better results than other approaches. The approach is able to remove horizontal and vertical lines and preserve the text in the sample.

C. PROPOSED METHOD ANALYSIS

As discussed with respect to our motivation, the proposed network structure uses wavelet-based multichannel images as the network input. Additionally, we present our method results with respect to the final performance. However, in this section, we explore and compare our proposed approach with some scenarios that can be considered.

1) IMPACT OF EACH STREAM AND DIFFERENT CHOICES

Although both multiple and single streams have the same information, multiple networks produce the best results among the three networks tested in our experiments. As shown in the results (Figs. 8 to 14), high-level and low-level features in a single network do not satisfy a robust binarization to obtain the high visual quality of foreground sections and a clean background simultaneously. Therefore, this structure is less effective than that of the multiple networks in terms of noise removal. Considering multiple network leads, each stream of the CNN separates the criterion of the binarization with the aim of obtaining better results. In addition, to demonstrate the optimal approach, we investigated the impact of each of the three networks trained by three multichannel image sets. We consider the multiple network with two streams and one stream that use two-channel images as input. Since it is possible to test and integrate the two streams of the two-channel images, we consider all three combinations ($C(3,2) = \frac{3!}{2!1!} = 3$) that can occur for the three sets of two-channel images (C_1, C_2 and C_3). The combinations can be considered for the first and second streams, integration of C_1 and C_2 (state 1), integration of C_1 and C_3 (state 2), and integration of C_2 and C_1 (state 3). Finally, we report the results for each of the two-channel images in one stream, separately. The experiments are tested on the DIBCO 2013 database, and based on SegNet, as shown in Table 6. We can see that our approach with three streams (integration of C_1, C_2 and C_3) achieves the best performance on DIBCO 2013. Additionally, it is observed that the integration of two streams is better than when we use one stream.

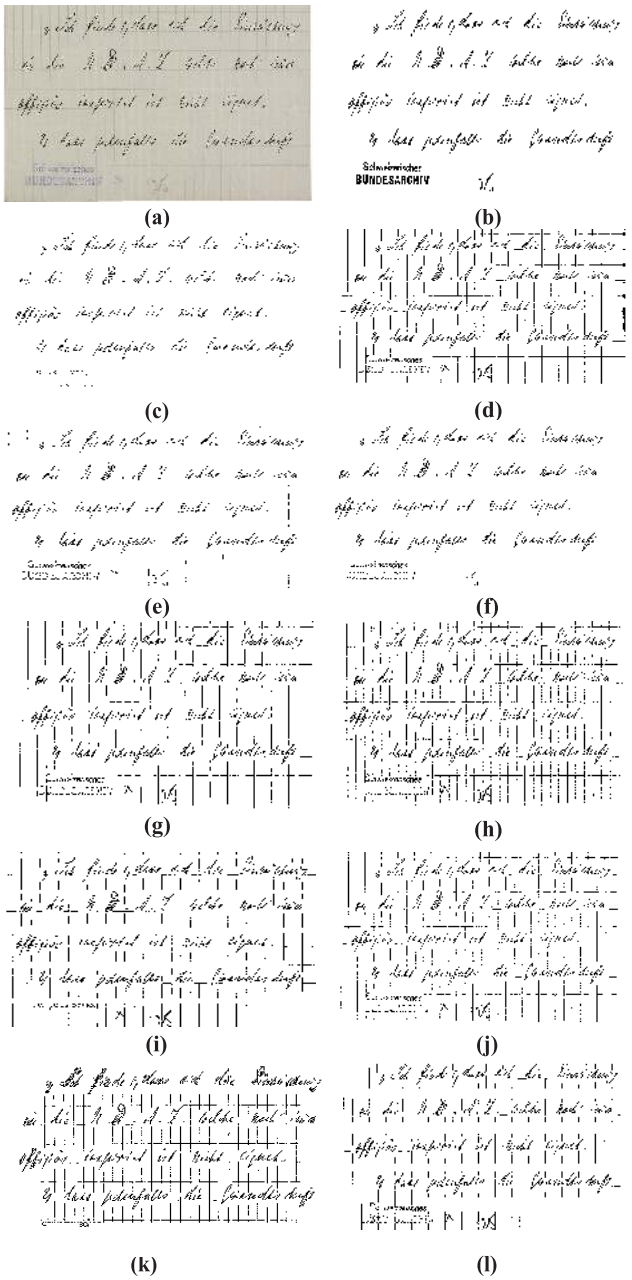

FIGURE 14. Binarization outputs for the challenging image (4) of the DIBCO 2019 database: (a) original image, (b) ground truth, (c) Otsu [13], (d) DeepLabv3+ (with original images), (e) our method (single network), (f) our method (multiple networks), (g) U-net (with original images), (h) our method (single network), (i) our method (multiple networks), (j) SegNet (with original images), (k) our method (single network), and (l) our method (multiple networks).

2) MIN AND MAX FUNCTIONS

In our approach for multiple networks, prediction maps are assembled by the minimum function. Using the prediction,

TABLE 6. Comparison with the binarization of different states of SegNet streams based on multichannel images, proceeded by the scores on the DIBCO 2013 image set.

States	FM (%)	F_{ps} (%)	PSNR (%)	DRD (%)
State 1 (C_1 and C_2)	94.80	96.45	20.96	2.20
State 2 (C_1 and C_3)	95.00	96.90	21.60	2.10
State 3 (C_2 and C_3)	94.70	96.15	21.85	1.97
C_1	94.50	96.00	20.80	1.86
C_2	94.90	95.95	21.40	2.08
C_3	94.55	96.05	21.67	2.07
C_1, C_2 and C_3	95.43	97.02	22.33	1.46

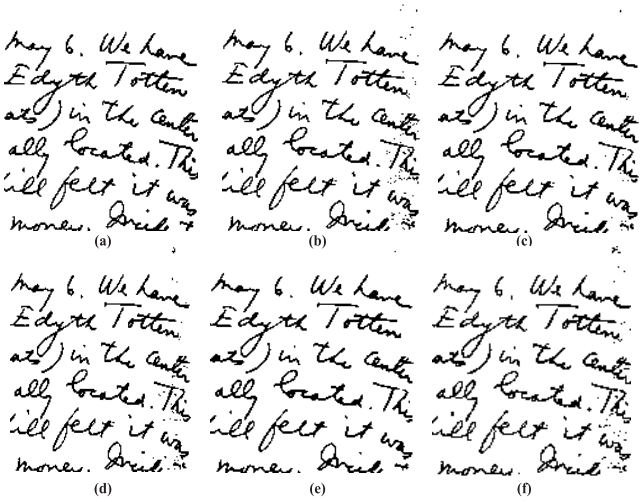

FIGURE 15. Sample image showing the effectiveness of the Min function based on HW1: (a) ground truth, (b) output of C_1, (c) output of C_2, (d) output of C_3, (e) integrated image (Min function) of the three networks (streams) and (f) integrated image (Max function) of the three networks (streams).

TABLE 7. Impact of Min and Max functions based on SegNet using the ICDAR 2011, ICDAR 2013 and ICFHR 2014 databases.

Functions	ICDAR 2011 (%)	ICDAR 2013 (%)	ICFHR 2014 (%)
Min	93.06	95.43	96.95
Max	92.15	94.80	94.98

this method will restrain the high probability value of noisy background pixels and blurring pixels along the contour of characters. This function can be used as a noise pixel removal tool with a high-probability value [5].

Another possible choice is the max function. Table 7 displays the evaluation of the max function on three databases. We can see that different F-measure values are obtained in terms of the three databases. Additionally, Fig. 15 shows the results for the three networks together with the result of integrating them in terms of Min and Max functions for three streams (C_1, C_2 and C_3). As shown, the integration of the three networks obtains an image with high visual quality when the Min function is used in the image.

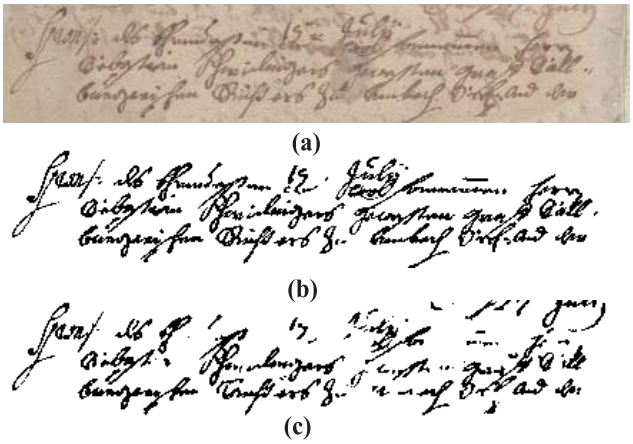

FIGURE 16. Binarization outputs of the challenging image (8) on the DIBCO 2016 database (failure case): (a) original image, (b) ground truth, and (c) our method (multiple networks).

3) REMAINING CHALLENGES

Although our method addresses certain challenges in the paper, additional efforts are required to improve the performance, mainly with respect to the handwritten documents. As shown in Fig. 16, the remaining main challenge occurs when the three problems of changing intensity in ink, ink stain, and similarity between text and background/foreground in an image occur concurrently. For this case, other networks can be tested on multichannel images, and multichannel images based on other filters are suggested.

VI. CONCLUSION

This investigation presents convolutional neural networks (CNNs) based upon multichannel images as input using wavelet analysis. Two approaches were considered to investigate the performance of wavelet-based multichannel images in document binarization. In the first method, a single network stream was trained using images with four channels, and in the other approach, multiple networks were trained by three sets of images with two channels. Our method from wavelet-based multichannel samples is considered. An evaluation of the effects of various quantifications on the seven chosen databases shows that our approach entirely exceeds the performance of SegNet, U-net and DeepLabv3+ without the use of multichannel images, and if we use multiple networks, the outputs are evidently better than those of the single-stream network. On three public databases, the recommended techniques exceed the performance of cutting edge methods. In total, our method somewhat addresses two weaknesses of text, namely, the similarity between the background and foreground and the oversight of certain strokes after binarization for selected weak strokes.

In subsequent works, our model may be improved to address the quandaries that arise when the existing problems of differing ink intensity, ink stain, and similarity between text and background occur simultaneously. The current network structure can be improved to solve this problem. The

use of new multichannel images with other filters in training might offer another solution. The advantage of our recommended approach proceeds beyond historical document application and may be used to tackle other domains, such as semantic segmentation for more than two classes and general classification issues.

ACKNOWLEDGMENT

The statements made herein are solely the responsibility of the authors.

REFERENCES

[1] S. Shariatmadari, S. Emadi, and Y. Akbari, "Nonlinear dynamics tools for offline signature verification using one-class Gaussian process," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 34, no. 01, Jan. 2020, Art. no. 2053001.

[2] Y. Akbari, M. J. Jalili, J. Sadri, K. Nouri, I. Siddiqi, and C. Djeddi, "A novel database for automatic processing of persian handwritten bank checks," *Pattern Recognit.*, vol. 74, pp. 253–265, Feb. 2018.

[3] Y. Akbari, K. Nouri, J. Sadri, C. Djeddi, and I. Siddiqi, "Wavelet-based gender detection on off-line handwritten documents using probabilistic finite state automata," *Image Vis. Comput.*, vol. 59, pp. 17–30, Mar. 2017.

[4] B. Gatos, K. Ntirogiannis, and I. Pratikakis, "ICDAR 2009 document image binarization contest (DIBCO 2009)," in *Proc. 10th Int. Conf. Document Anal. Recognit.*, Jul. 2009, pp. 1375–1382.

[5] Q. N. Vo, S. H. Kim, H. J. Yang, and G. Lee, "Binarization of degraded document images based on hierarchical deep supervised network," *Pattern Recognit.*, vol. 74, pp. 568–586, Feb. 2018.

[6] T. Wang, W. Sun, H. Qi, and P. Ren, "Aerial image super resolution via wavelet multiscale convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 769–773, May 2018.

[7] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Process.*, vol. 1, no. 2, pp. 205–220, Apr. 1992.

[8] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.

[9] R. Kemker, C. Salvaggio, and C. Kanan, "Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 60–77, Nov. 2018.

[10] C. Tensmeyer and T. Martinez, "Document image binarization with fully convolutional neural networks," 2017, *arXiv:1708.03276*. [Online]. Available: http://arxiv.org/abs/1708.03276

[11] C. Tensmeyer and T. Martinez, "Historical document image binarization: A review," *Social Netw. Comput. Sci.*, vol. 1, no. 3, p. 26, May 2020.

[12] Sulaiman, Omar, and Nasrudin, "Degraded historical document binarization: A review on issues, challenges, techniques, and future directions," *J. Imag.*, vol. 5, no. 4, p. 48, Apr. 2019.

[13] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.

[14] W. Niblack, *An Introduction to Digital Image Processing*. København, Denmark: Strandberg Publishing Company, 1985.

[15] J. Sauvola and M. Pietikäinen, "Adaptive document image binarization," *Pattern Recognit.*, vol. 33, no. 2, pp. 225–236, Feb. 2000.

[16] K. V. Mardia and T. J. Hainsworth, "A spatial thresholding method for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 10, no. 6, pp. 919–927, 1988.

[17] Z. Hadjadj, M. Cheriet, A. Meziane, and Y. Cherfa, "A new efficient binarization method: Application to degraded historical document images," *Signal, Image Video Process.*, vol. 11, no. 6, pp. 1155–1162, Sep. 2017.

[18] S. Lu, B. Su, and C. L. Tan, "Document image binarization using background estimation and stroke edges," *Int. J. Document Anal. Recognit.*, vol. 13, no. 4, pp. 303–314, Dec. 2010.

[19] S. Pardhi and G. U. Kharat, "An improved binarization method for degraded document," *Int. J. Res. Advent Technol.*, pp. 1–5, 2017.

[20] C.-H. Chou, W.-H. Lin, and F. Chang, "A binarization method with learning-built rules for document images produced by cameras," *Pattern Recognit.*, vol. 43, no. 4, pp. 1518–1530, Apr. 2010.

[21] B. Gatos, I. Pratikakis, and S. J. Perantonis, "Improved document image binarization by using a combination of multiple binarization techniques and adapted edge information," in *Proc. 19th Int. Conf. Pattern Recognit.*, Dec. 2008, pp. 1–4.

[22] K. Ntirogiannis, B. Gatos, and I. Pratikakis, "A combined approach for the binarization of handwritten document images," *Pattern Recognit. Lett.*, vol. 35, pp. 3–15, Jan. 2014.

[23] N. R. Howe, "A Laplacian energy for document binarization," in *Proc. Int. Conf. Document Anal. Recognit.*, Sep. 2011, pp. 6–10.

[24] N. R. Howe, "Document binarization with automatic parameter tuning," *Int. J. Document Anal. Recognit.*, vol. 16, no. 3, pp. 247–258, Sep. 2013.

[25] A. Mishra, K. Alahari, and C. V. Jawahar, "Unsupervised refinement of color and stroke features for text binarization," *Int. J. Document Anal. Recognit.*, vol. 20, no. 2, pp. 105–121, Jun. 2017.

[26] Y. Chen and L. Wang, "Broken and degraded document images binarization," *Neurocomputing*, vol. 237, pp. 272–280, May 2017.

[27] F. Jia, C. Shi, K. He, C. Wang, and B. Xiao, "Degraded document image binarization using structural symmetry of strokes," *Pattern Recognit.*, vol. 74, pp. 225–240, Feb. 2018.

[28] K. R. Ayyalasomayajula and A. Brun, "Historical document binarization combining semantic labeling and graph cuts," in *Proc. Scand. Conf. Image Anal.* Springer, 2017, pp. 386–396.

[29] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[30] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, Sep. 2004.

[31] G. Meng, K. Yuan, Y. Wu, S. Xiang, and C. Pan, "Deep networks for degraded document image binarization through pyramid reconstruction," in *Proc. 14th IAPR Int. Conf. Document Anal. Recognit. (ICDAR)*, Nov. 2017, pp. 727–732.

[32] X. Peng, H. Cao, and P. Natarajan, "Using convolutional encoder-decoder for document image binarization," in *Proc. 14th IAPR Int. Conf. Document Anal. Recognit. (ICDAR)*, Nov. 2017, pp. 708–713.

[33] F. Westphal, N. Lavesson, and H. Grahn, "Document image binarization using recurrent neural networks," in *Proc. 13th IAPR Int. Workshop Document Anal. Syst. (DAS)*, Apr. 2018, pp. 263–268.

[34] J. Calvo-Zaragoza and A.-J. Gallego, "A selectional auto-encoder approach for document image binarization," *Pattern Recognit.*, vol. 86, pp. 37–47, Feb. 2019.

[35] J. Pastor-Pellicer, S. España-Boquera, F. Zamora-Martínez, M. Z. Afzal, and M. J. Castro-Bleda, "Insights on the use of convolutional neural networks for document image binarization," in *Proc. Int. Work-Conf. Artif. Neural Netw.* Springer, 2015, pp. 115–126.

[36] M. Z. Afzal, J. Pastor-Pellicer, F. Shafait, T. M. Breuel, A. Dengel, and M. Liwicki, "Document image binarization using LSTM: A sequence learning approach," in *Proc. 3rd Int. Workshop Historical Document Imag. Process. (HIP)*, 2015, pp. 79–84.

[37] K. R. Ayyalasomayajula, F. Malmberg, and A. Brun, "PDNet: Semantic segmentation integrated with a primal-dual network for document binarization," *Pattern Recognit. Lett.*, vol. 121, pp. 52–60, Apr. 2019.

[38] Y. Akbari, A. S. Britto, S. Al-maadeed, and L. S. Oliveira, "Binarization of degraded document images using convolutional neural networks based on predicted two-channel images," in *Proc. Int. Conf. Document Anal. Recognit. (ICDAR)*, Sep. 2019, pp. 973–978.

[39] S. Kang, B. K. Iwana, and S. Uchida, "Cascading modular U-Nets for document image binarization," in *Proc. Int. Conf. Document Anal. Recognit. (ICDAR)*, Sep. 2019, pp. 675–680.

[40] I. Pratikakis, B. Gatos, and K. Ntirogiannis, "H-DIBCO 2010-handwritten document image binarization competition," in *Proc. 12th Int. Conf. Frontiers Handwriting Recognit.*, Nov. 2010, pp. 727–732.

[41] I. Pratikakis, B. Gatos, and K. Ntirogiannis, "ICDAR 2011 document image binarization contest (DIBCO 2011)," in *Proc. Int. Conf. Document Anal. Recognit.*, Sep. 2011, pp. 1506–1510.

[42] I. Pratikakis, B. Gatos, and K. Ntirogiannis, "ICFHR 2012 competition on handwritten document image binarization (H-DIBCO 2012)," in *Proc. Int. Conf. Frontiers Handwriting Recognit.*, Sep. 2012, pp. 817–822.

[43] I. Pratikakis, B. Gatos, and K. Ntirogiannis, "ICDAR 2013 document image binarization contest (DIBCO 2013)," in *Proc. 12th Int. Conf. Document Anal. Recognit.*, Aug. 2013, pp. 1471–1476.

[44] K. Ntirogiannis, B. Gatos, and I. Pratikakis, "ICFHR2014 competition on handwritten document image binarization (H-DIBCO 2014)," in *Proc. 14th Int. Conf. Frontiers Handwriting Recognit.*, Sep. 2014, pp. 809–813.

[45] R. G. Mesquita, C. A. B. Mello, and L. H. E. V. Almeida, "A new thresholding algorithm for document images based on the perception of objects by distance," *Integr. Comput.-Aided Eng.*, vol. 21, no. 2, pp. 133–146, Mar. 2014.
[46] I. Pratikakis, K. Zagoris, G. Barlas, and B. Gatos, "ICFHR2016 handwritten document image binarization contest (H-DIBCO 2016)," in *Proc. 15th Int. Conf. Frontiers Handwriting Recognit. (ICFHR)*, Oct. 2016, pp. 619–623.
[47] S. Katz, A. Tal, and R. Basri, "Direct visibility of point sets," *ACM Trans. Graph. (TOG)*, vol. 26, no. 3, p. 24, 2007.
[48] I. Pratikakis, K. Zagoris, G. Barlas, and B. Gatos, "ICDAR2017 competition on document image binarization (DIBCO 2017)," in *Proc. 14th IAPR Int. Conf. Document Anal. Recognit. (ICDAR)*, vol. 1, Nov. 2017, pp. 1395–1403.
[49] O. Ronneberger, P. Philipp, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Springer, 2015, pp. 234–241.
[50] I. Pratikakis, K. Zagori, P. Kaddas, and B. Gatos, "ICFHR 2018 competition on handwritten document image binarization (H-DIBCO 2018)," in *Proc. 16th Int. Conf. Frontiers Handwriting Recognit. (ICFHR)*, Niagara Falls, NY, USA, Aug. 2018, pp. 489–493.
[51] I. Pratikakis, K. Zagoris, X. Karagiannis, L. Tsochatzidis, T. Mondal, and I. Marthot-Santaniello, "ICDAR 2019 Competition on Document Image Binarization (DIBCO 2019)," in *Proc. Int. Conf. Document Anal. Recognit. (ICDAR)*, vol. 1. Sep. 2019, pp. 1547–1556.
[52] K. Ntirogiannis, B. Gatos, and I. Pratikakis, "Performance evaluation methodology for historical document image binarization," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 595–609, Feb. 2013.
[53] X. Wang, X. Ou, B.-W. Chen, and M. Kim, "Image denoising based on improved wavelet threshold function for wireless camera networks and transmissions," *Int. J. Distrib. Sensor Netw.*, vol. 11, no. 9, Sep. 2015, Art. no. 670216.
[54] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
[55] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556
[56] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 801–818.
[57] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*. [Online]. Available: http://arxiv.org/abs/1706.05587
[58] H. Z. Nafchi, S. M. Ayatollahi, R. F. Moghaddam, and M. Cheriet, "An efficient ground truthing tool for binarization of historical manuscripts," in *Proc. 12th Int. Conf. Document Anal. Recognit.*, Aug. 2013, pp. 807–811.
[59] F. Deng, Z. Wu, Z. Lu, and M. S. Brown, "BinarizationShop: A user-assisted software suite for converting old documents to black-and-white," in *Proc. 10th Annu. Joint Conf. Digit. Libraries*, 2010, pp. 255–258.
[60] R. Hedjam, H. Z. Nafchi, R. F. Moghaddam, M. Kalacska, and M. Cheriet, "ICDAR 2015 contest on MultiSpectral text extraction (MS-TEx 2015)," in *Proc. 13th Int. Conf. Document Anal. Recognit. (ICDAR)*, Aug. 2015, pp. 1181–1185.
[61] L. Yue, H. Shen, J. Li, Q. Yuan, H. Zhang, and L. Zhang, "Image super-resolution: The techniques, applications, and future," *Signal Process.*, vol. 128, pp. 389–408, Nov. 2016.
[62] C. Lim Tan, R. Cao, and P. Shen, "Restoration of archival documents using a wavelet technique," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 10, pp. 1399–1404, Oct. 2002.
[63] R. F. Moghaddam and M. Cheriet, "A variational approach to degraded document enhancement," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 8, pp. 1347–1361, Aug. 2010.
[64] J. Chen, X. Kang, Y. Liu, and Z. J. Wang, "Median filtering forensics based on convolutional neural networks," *IEEE Signal Process. Lett.*, vol. 22, no. 11, pp. 1849–1853, Nov. 2015.
[65] Y. Duan, F. Liu, L. Jiao, P. Zhao, and L. Zhang, "SAR image segmentation based on convolutional-wavelet neural network and Markov random field," *Pattern Recognit.*, vol. 64, pp. 255–267, Apr. 2017.
[66] B. Gatos, I. Pratikakis, and S. J. Perantonis, "An adaptive binarization technique for low quality historical documents," in *Proc. Int. Workshop Document Anal. Syst.* Springer, 2004, pp. 102–113.
[67] B. Su, S. Lu, and C. Lim Tan, "Robust document image binarization technique for degraded document images," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1408–1417, Apr. 2013.
[68] T. Lelore and F. Bouchara, "FAIR: A fast algorithm for document image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 2039–2048, Aug. 2013.

YOUNES AKBARI received the B.Sc. degree in computer software engineering from Payame Noor University (Central Branch of Birjand), Iran, in 2006, the M.Sc. degree in information technology management from Payame Noor University (Central Branch of Tehran), Tehran, Iran, in 2011, and the Ph.D. degree in applied mathematics (numerical analysis) from the University of Semnan, Iran, in 2017. He was a Postdoctoral Researcher with the Department of Computer Science and Engineering, Qatar University (QU), and worked on the project of document binarization. He is a Reviewer for several international journals in the field of artificial intelligence, such as the IEEE TRANSACTIONS ON CYBERNETICS, *Pattern Recognition*, and *Artificial Intelligence Review*. His research interests include pattern recognition, neural networks, remote sensing, and document analysis. He has published several articles on these areas.

SOMAYA AL-MAADEED (Senior Member, IEEE) received the Ph.D. degree in computer science from Nottingham, U.K., in 2004. She was a Visiting Academician with Northumbria University, U.K. She is currently the Head of the Department of Computer Science, Qatar University. She is also the Coordinator of the Computer Vision Research Group, Qatar University. She organized several workshops and competitions related to computer vision. From 2012 to 2013, she was selected as a Participant of the Current and Future Executive Leaders Program, Qatar Leadership Centre, which was established in 2008 by an Emiri Decree. She has supervised students through research projects related to community and industry. She enjoys excellent collaboration with industry and national and international institutions. She is also a Principal Investigator of several funded research projects generating approximately five million dollars in the last years. She has published extensively in computer vision and pattern recognition and delivered workshops on teaching programming for undergraduate students. She attended workshops related to higher education strategy, assessment methods, and interactive teaching. In 2015, she was elected as the IEEE Chair of the Qatar Section. She and her team were one of the recipients of the Best Performance Awards at ICDAR 2011 and ICDAR 2015 Signature Verification.

KALTHOUM ADAM (Member, IEEE) received the B.Sc. degree in computer science from UAE University, and the M.Sc. degree in computer science from SDSU, San Diego, CA, USA. She is currently pursuing the Ph.D. degree in computer science with Qatar University. She worked as an IT Auditor at CHASE, USA. She also worked as an IT Instructor with the Ministry of Education, UAE. Her main research interests include machine learning and image processing.