IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received February 22, 2021, accepted February 25, 2021, date of publication March 1, 2021, date of current version March 9, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3062904

Document Image Binarization With Stroke Boundary Feature Guided Network

QUANG-VINH DANG AND GUEE-SANG LEE, (Member, IEEE)
Department of Artificial Intelligence Convergence, Chonnam National University, Gwangju 61186, South Korea

Corresponding author: Guee-Sang Lee (gslee@jnu.ac.kr)

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant NRF-2018R1D1A3B05049058 and Grant NRF-2020R1A4A1019191.

ABSTRACT Text is the most crucial element in a document image but is often disconnected in document image binarization. Most of the previous methods based on deep learning do not focus on structure information such as stroke boundary, leading to disconnected strokes when the stroke is ambiguous or weak. In this paper, we propose a multi-task learning with an auxiliary task for learning stroke boundary features in an adversarial manner. The learned boundary features are integrated into the main task for the binarization. Specifically, in the first step, in addition to using shared global location features with the main task, the auxiliary task leverages additional local edges to obtain stroke boundary features. In the second step, we use adversarial loss based on boundary ground truth to supervise the obtained stroke boundary feature in the auxiliary task. The adversarial training is to embed expert knowledge, especially structure information, in the model. In the third step, the learned boundary feature from the auxiliary task supports the main task directly. The fusion module of the main task refines the final binarized image. Experiments show that our method achieves better-preserved stroke and better performance than existing methods on benchmark H-DBCO and DIBCO datasets.

INDEX TERMS Document image binarization, auxiliary-task learning, stroke boundary feature, multi-task learning, preserved stroke.

I. INTRODUCTION

Binarization of degraded document images is to assign each pixel for one of the two following classes: foreground and background. It plays a significant role in document image analysis domains such as optical character recognition, content understanding, and document layout recognition. History document images suffer from different degradations such as ink stains, noise, spots, smears, low-contrast ink strokes, and so forth. This is because storage conditions and maintenance are insufficient. Although binarization is relatively straightforward for uniform images, it is quite difficult for degraded document images. This importance leads to hold relating competitions such as document image binarization (DIBCO) [1] that started in 2009. Therefore, it is essential to study the binarization for degraded document images.

Over the past few years, computer vision methods applying deep learning have achieved great achievements, surpassing conventional methods such as recognition, classification, and segmentation. The document image binarization has impressive results in applying deep learning. Many methods have been advent to binarizing document images, especially after the emerging of Fully Convolutional Neural Networks (FCNs) [2]. They are gradually substituting traditional image binarization methods because of their outstanding ability to learn data distribution. Recently, there are many developed frameworks to solve the problem in the document image binarization task as in [3]–[5]. The papers' results have efficiency and high performance. However, weak or ambiguous strokes are often disconnected after the binarization process. While in [6], a hierarchical deep supervised network architecture extracts different feature levels for predicting the text pixels. Despite these efforts, developing preserved strokes, especially for ambiguous or thin structures in document images, is still challenging. These approaches classify per pixel from the image region without focusing on the structure information, supporting text-stroke preservation effectively.

To tackle the aforementioned problems, we propose a stroke boundary feature guided framework based on multi-task learning [7]. It achieves better-preserved stroke

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

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

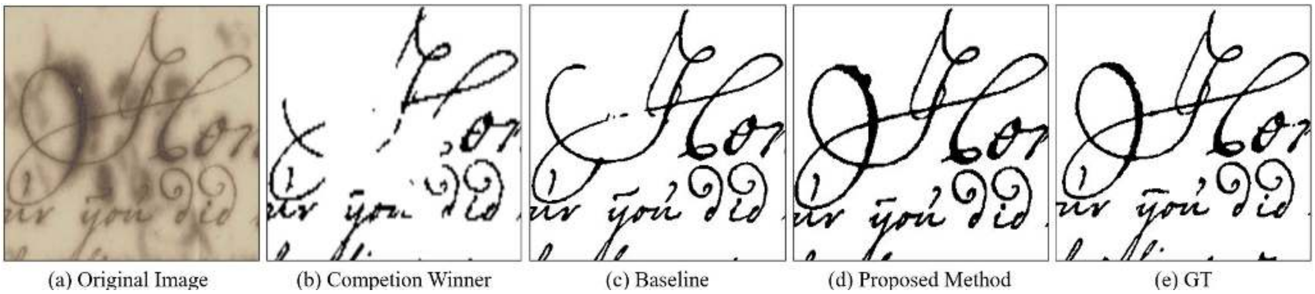

FIGURE 1. Results of binarization methods on degraded document image with the ambiguous strokes. (a) Shows the degraded document image. (b) Illustrate the results of [8]. (c) Is the result of baseline and our proposed method is displayed in (d). (e) Is groundtruth. Ours has better-preserved stroke than other methods.

and better performance than existing methods (see Fig.1). We design a stroke boundary feature extracting module that exploits both additional local edge features and shared global location features. Then, we incorporate the module into the auxiliary task to predict the text-stroke structure in the form of stroke boundary maps. The auxiliary task is trained in an adversarial manner with a discriminator network to embed expert knowledge in the model. We design the feature fusion module to combine the obtained stroke boundary feature from the auxiliary task and stroke feature from the main task to have final better-preserved stroke results. In summary, this paper makes four significant contributions:

- We propose a stroke boundary feature guided framework. Auxiliary task learns text-stroke structure information in the form of stroke boundary to support the main task, leading to degraded document image binarization results having better-preserved strokes than existing methods.
- We design the stroke boundary feature extracting module, which jointly exploits both additional local edge features and shared global location features to obtain stroke boundary features in the auxiliary task. To embed expert knowledge, especially structure information, in the model, the auxiliary task is trained in an adversarial manner with the discriminator network.
- We also design the feature fusion module to fuse both the obtained stroke boundary feature from the auxiliary task and stroke feature from the main task to have final better-preserved stroke results.
- The proposed method achieves better results than existing methods on the H-DIBCO and DIBCO datasets.

The rest of the paper is organized as follows. In section II, related works are mentioned. Section III describes the proposed method for document image binarization. Experimental results are discussed in section IV. Finally, the paper is concluded in section V.

II. RELATED WORKS

Conventional binarization methods are based on statistical knowledge. It is a nonparametric and unsupervised method of automatic threshold selection for document image binarization. It employs global, local, or hybrid thresholds to classify each pixel of an image as either text or background. In terms of the global threshold method, the pixels of the whole image are classified by a fixed threshold, such as Otsu's [9], Kittler's [10] and Brink's [11] methods. It automatically estimates a value according to the input image. Specifically, it maximizes the distance between background and text. Then, it finds the maximum interclass variation to binarize degraded document images. The global thresholding approach operates very well on uniform and clean images. However, it generates bad results on degraded document images with non-uniform backgrounds. With the local thresholding method, it estimates local intensity to classify each pixel, such as Niblack [12] and Sauvola [13]. In other words, they consider chosen background pixels based on each neighborhood locally. About hybrid thresholding method, it combines both local and global thresholds. In [14], the authors used polynomial smoothing to classify the background. Then, the local threshold is applied to capture the foreground text. In [15], authors split input images into blocks. Different binarization methods binarize the blocks to improve performance. Recently, in [16], authors proposed the selective diffusion model that involves reaction for binarizing bleed-through document images. It utilizes the Perona-Malik diffusion to selectively smooth document images, leading to remove bleed-through. Besides, the nonlinear reaction term is responsible for the desired binarization.

In addition, some document binarization methods utilized priori knowledge such as edge information to support for binarizing process. Specifically, in [17], authors combined the global thresholding methods and the edge detecting method. However, it only integrate the edge information into global thresholding methods without focusing on the intensity of non-uniform objects. In [18], the Canny edge detector [19] is applied to detect edge pixels in original images. After the closed image edges are considered as seeds to find the text region, the transition pixels are computed based on the intensity differences in small neighbor regions and the edge pixels. Finally, computed threshold is derived from the statistical information of these pixels. In [20], the authors also propose the automatic parameter tuning model along with the edge information based on the canny detector to generate a better binarization result. Conventional binarization methods

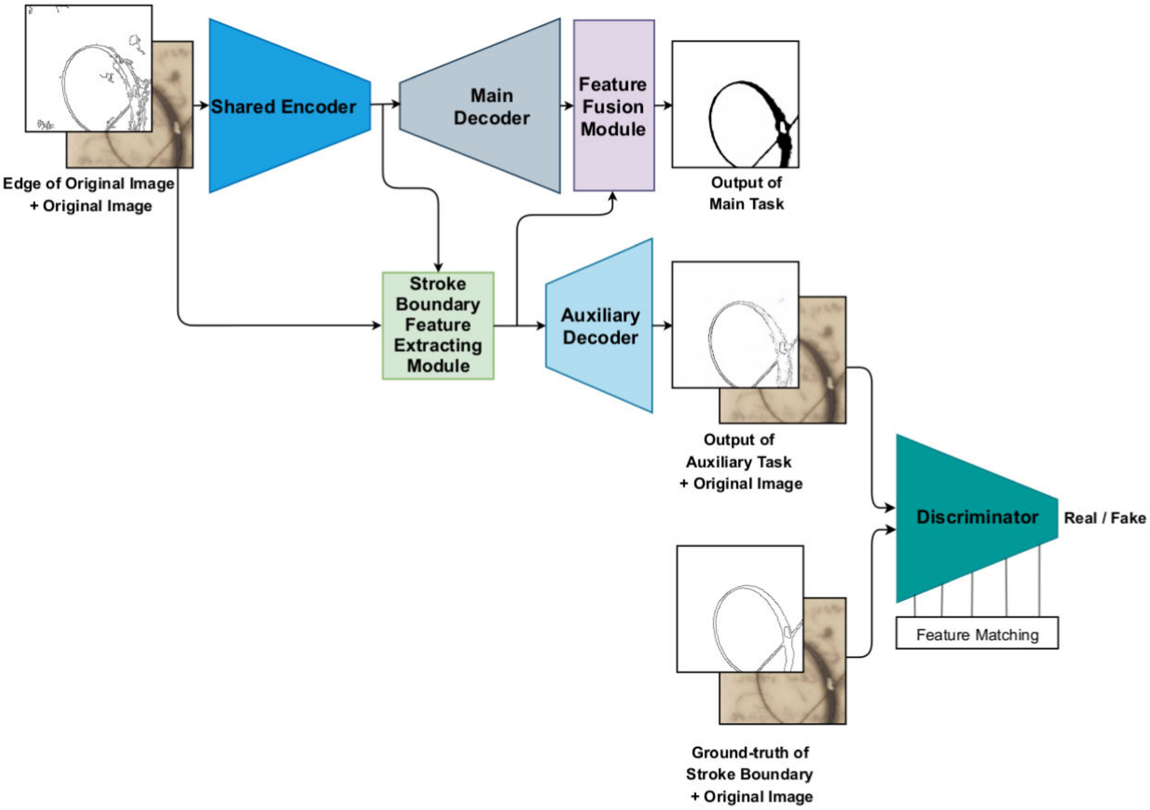

FIGURE 2. Summary of the proposed method. We apply Canny edge detector [19] to detect edge from original image. Edge of original image and original image are the input of network. Output of main task is binarized images. Auxiliary task is learned by adversarial training. Output of auxiliary task is the boundary of stroke.

work well on images having simple backgrounds. However, they produce poor results on images having complex backgrounds.

With the emerging powerful deep convolutional neural network, many developed techniques achieve good performance on various applications such as object classification, segmentation, and detection. Besides, many methods for document image binarization are also established based on deep learning as in [6], [21], [22]. In [21], the authors introduce binarization task as each pixel classification. They applied a fully convolutional network with multiple image scales. In [6], a hierarchical deep supervised network (DSN) model predicts text or background pixels at various feature levels. With higher-level features, text pixels are distinguished significantly from background noises. With lower-level features, predicted text pixels show sharply at the boundary area. Then, they combined these features to produce better results. In [22], the authors utilize convolutional auto-encoders to learn how to map an input image to its output, in which activations specify pixels belonging to either foreground or background. Consequently, the trained model can analyze degraded document images. Then, they apply a global threshold to binarize the final feature map. In [4], the DeepOtsu network enhances the degraded document images by refining the outputs iteratively. Then, the authors apply a global or local threshold to the binarization map. Recently, in [23], to solve the problem of a shortage of training images in DIBCO datasets, the authors proposed Cascading Modular U-Nets. Specifically, each modular U-Net is trained on a large generic image dataset coco-text with specific tasks. Then, the pre-trained U-Nets modules are cascaded with the inter-module skip-connections for enhancing the connectivity between the modules. Finally, the authors conduct a fine-tuning process on DIBCO datasets for improving overall cascading model performance. Although these methods achieved good performance, they don't exploit structure information completely to support image binarization processing. Therefore, in this paper, we propose an auxiliary task to learn structure information such as stroke boundary features in an adversarial manner and integrate the obtained boundary features into the main task for document image binarization.

III. PROPOSED METHOD

The framework of the proposed model is based on multi-task learning [7], including the main task and auxiliary task. We design the stroke boundary feature extracting module for preserving better stroke edge information in the auxiliary task. The designed feature fusion module fuses these stroke boundary features and multi-level stroke features derived from the main decoder in the main task to obtain better-preserved stroke and better performance as in Fig. 2.

A. SHARED ENCODER

We design the shared encoder as the encoder part in BasNet [24]. However, to increase edge information in terms of stroke boundary preserving benefits, the proposed model's input is original images conditioned on their edge images that are detected by canny detector [19]. Therefore, the input convolution layer is modified with convolution filters having four input channels. Furthermore, to further capture global information, we add the encoder's top that consists of one convolution layer with 3×3 filters having dilation [25] equal to 1, followed by two convolution layers with 3×3 filters having dilation equal to 2. A batch normalization [26] and a ReLU activation function [27] are applied to each of these convolution layers. The extracted features from this encoder are shared between the main task and the auxiliary task.

B. MAIN TASK

1) MULTI-LEVEL STROKE INFORMATION

The main decoder is almost symmetrical to the shared encoder. The main decoder includes six stages. Each stage comprises three convolution layers, followed by batch normalization and ReLU activation function. Each stage's input is the features that are concatenated the upsampled output from its previous stage and its corresponding stage output in the encoder. The outputs of each stage have multi-level stroke information. As the convolutional features at different levels identify the object and its surroundings from different views [28], we employ them to feed into the feature fusion module for simultaneously incorporating the different coarse semantics and the fine details from stroke boundary feature.

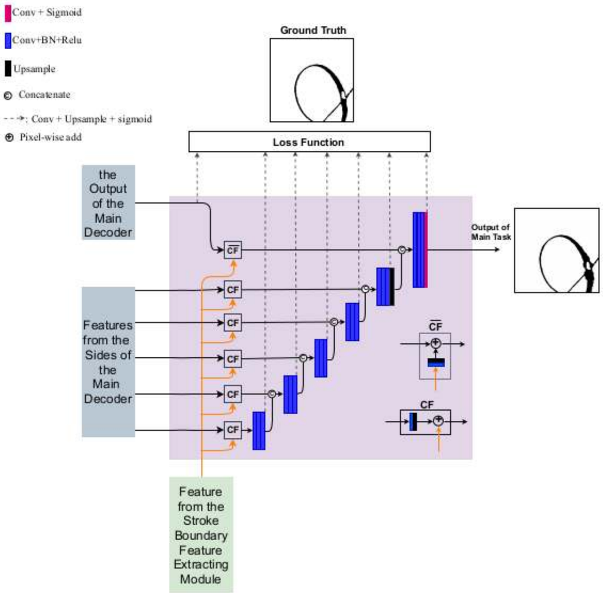
Conv + Sigmoid
Conv+BN+Relu
Upsample
Concatenate
- - →: Conv + Upsample + sigmoid
Pixel-wise add
Ground Truth
Loss Function
the Output of the Main Decoder
Features from the Sides of the Main Decoder
Feature from the Stroke Boundary Feature Extracting Module
$\overline{CF}$
CF
CF
CF
CF
CF
Output of Main Task
$\overline{CF}$
CF

FIGURE 3. Architecture of the feature fusion module.

2) FEATURE FUSION MODULE

We aim to utilize the stroke boundary features to guide each of the multi-level stroke features to have a better prediction in preserving strokes and performance. We fuse the enhanced stroke boundary features and the enhanced multi-level stroke features to use complementary information by deep supervision on side responses [29]. Hence, we propose the feature fusion module as Fig. 3.

Specifically, the module's input is multi-level stroke features mentioned in Section III-B1 and stroke boundary features mentioned in Section III-C1. We apply convolution on each side features from the main decoder, followed by batch normalization, Relu activation function, and unsampling them. It aims to change the channel and resolution of each side feature as stroke boundary features. Then, we add stroke boundary features to each of these adjusted side features, as shown in the CF (combination of features) block of Fig. 3.

As the channel number of the stroke boundary features is more than the main decoder's output, and the resolution of the stroke boundary is smaller than the output of the main decoder, we apply convolution on the stroke boundary feature, followed by batch normalization, Relu activation function and unsampling them. It aims to adjust the channel and resolution of the stroke boundary features as the main decoder's output. After this, we add these modified stroke boundary features to the main decoder's output, as shown in the $\overline{CF}$ block of Fig. 3.

The feature fusion module includes six stages as the main decoder. Each stage includes three convolution layers, followed by batch normalization and a ReLU activation function. Each stage's input is the concatenated feature maps from its previous stage and its corresponding fused features from the CF block. The final stage input is the concatenated feature maps of the upsampled output from its previous stage and its corresponding fused features from $\overline{CF}$ block. The output of the final stage is the model's output.

C. AUXILIARY TASK

1) STROKE BOUNDARY FEATURE EXTRACTION

The low-level feature maps in layer 1 of VGG [30] have edge-preserving properties [31]–[33]. These low-level features can help to predict stroke boundary. Because layer 1 of VGG is too close to the input and the receptive field is too small, we employ feature maps in layer 2 for preserving better edge information, as mentioned in [33]. We can use low-level features of layer 2 simply from the encoder as in [33]. However, we expect that the stroke boundary feature extracting module focus on edge information more in the auxiliary task. We use pre-trained layers 1 and 2 of VGG for extracting edge features separately from the shared encoder. In order to get stroke boundary features, however, only edge information is not enough. It is also important to provide high-level semantic information or global location information.

The global location information gradually increases from the low-level layer to the top-level layer in the shared encoder. In other words, the top-level layer's receptive field is the largest, and the global location is the most accurate. We apply convolution on the global location feature, followed by the Relu activation function and unsampling them. It aims to adjust the channel and resolution of the global location features as edge features mentioned above. After this, we add these modified global location features to edge features. To obtain more robust stroke boundary features, we apply three convolutional layers, followed by the Relu activation function on the added features, as shown in Fig. 4. Finally, we obtain stroke boundary information, and it needs to be learned.

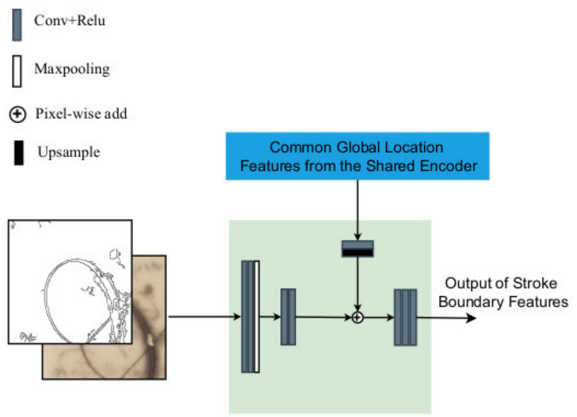
Conv+Relu
Maxpooling
Pixel-wise add
Upsample
Common Global Location Features from the Shared Encoder
Output of Stroke Boundary Features

FIGURE 4. Architecture of the stroke boundary feature extracting module.

We design the auxiliary decoder that unsample stroke boundary feature to feature map having the same resolution as the original images and single-channel. To extract the learned stroke boundary feature, we use the auxiliary decoder to convert the stroke boundary features to the probability map and train the auxiliary task in an adversarial way.

2) STROKE BOUNDARY INFORMATION-BASED DISCRIMINATION NETWORK

We expect that expert knowledge about stroke boundary information is embedded in the fully automatic image binarization model. To embed expert knowledge of structure boundary in the model, authors in [34] propose shape boundary-aware evaluator based on the discrimination network in an adversarial way without the user interaction. For this purpose, we propose a discrimination network based on structure stroke boundary information, as shown in Fig. 5. The discrimination network evaluates how much predicted stroke boundary features are preserved to the network for the auxiliary task by using the ground-truth stroke boundary. In particular, the discrimination network's input comprises the original image and stroke boundary image (predicted or the ground-truth). The original image and given stroke boundary image are concatenated and fed into the discrimination network. Then, the discrimination network is to evaluate whether the stroke boundary segmentation predictions are consistent with the original image or not. For example, the discrimination network provides a high evaluation score with the ground-truth stroke boundary map and the original image. In contrast, given the poorly predicted stroke boundary map and original image, the discrimination network provides a low evaluation score as the poorly predicted stroke boundary map is not consistent with the original image. We trained the network in an adversarial manner with the loss function mentioned in Section III-D2.

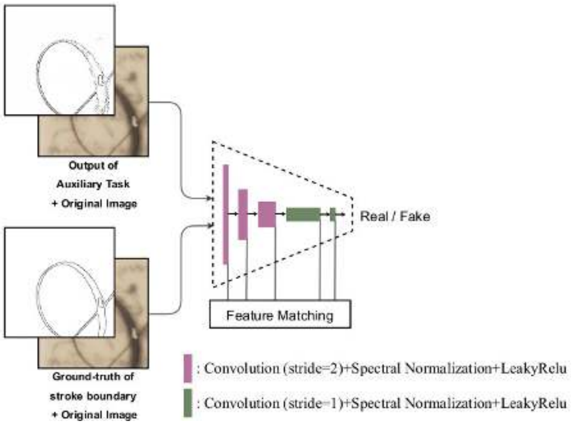
Output of Auxiliary Task + Original Image
Real / Fake
Feature Matching
Ground-truth of stroke boundary + Original Image
: Convolution (stride=2)+Spectral Normalization+LeakyRelu
: Convolution (stride=1)+Spectral Normalization+LeakyRelu

FIGURE 5. Architecture of Stroke Boundary Information-based Discrimination Network.

D. LOSS FUNCTION

In the proposed model, the input uses the degraded document image I_{org}, conditioned its edge map I_{edge_org} computed using Canny edge detector. The network G, including the main task G_{main} and auxiliary task G_{au}, generates binarized map O_{pre} and predicted stroke boundary map O_{bdr_pre}

$$O_{pred}, O_{bdr\_pre} = G(I_{org}, I_{edge\_org}) \tag{1}$$

To train the proposed model, including the main task and the auxiliary task in an adversarial manner, we utilize different types of loss corresponding to the probability feature maps for each task. The total loss function for training the network can be expressed as

$$Loss = Loss_{main\_task} + \alpha Loss_{auxiliary\_task} \tag{2}$$

where α is the weight of $Loss_{auxiliary\_task}$. For our experiments, we choose $\alpha = 0.5$

1) LOSS FUNCTION FOR MAIN TASK

In the main task, the input employs the degraded document image I_{org} and its edge map I_{edge_org}. The network for the main task G_{main} has the output being binarization map O_{pre}

$$O_{pre} = G_{main}(I_{org}, I_{edge\_org}) \tag{3}$$

Our training loss for the main task is computed as the summation over all side-outputs and main task output O_{pre}

$$Loss_{main\_task} = \sum_{i=1}^{I} l^i \tag{4}$$

where l^i is the loss of the i-th side output and main task output O_{pre}. I denotes the total number of the outputs. As shown in Fig 3, I is equal to 7, including six side outputs from the feature fusion module and one final output from the main task O_{pre}

We apply hybrid loss [24] for l^i

$$l^i = l_{bce} + l_{iou} + l_{ssim} \tag{5}$$

where $l_{bce}, l_{iou}, l_{ssim}$ define BCE loss [35], IOU loss [36], and SSIM loss [37], respectively. Each loss is calculated based on the output (each side output or main task output O_{pre}) and gound-truth binarized image.

2) LOSS FUNCTION FOR AUXILIARY TASK

In the auxiliary task, the input employs the degraded document image I_{org} and its edge map I_{edge_org}. The network for auxiliary task G_{au} predicts the stroke boundary map O_{bdr_pre}

$$O_{bdr\_pre} = G_{au}(I_{org}, I_{edge\_org}) \tag{6}$$

We use the ground-truth image of the stroke boundary I_{bdr_gt} and O_{bdr_pre} conditioned on I_{org} as inputs of the discriminator that predicts whether the stroke boundary map is real or not. The network for auxiliary task is trained with an objective combined of the hinge variant of GAN loss [38] and feature-matching loss [39].

$$Loss_{auxiliary\_task} = \alpha_{au} l_{G_{au}} + \alpha_{fm} l_{fm} \tag{7}$$

where α_{au} and α_{fm} are regularization parameters. For our experiment, we choose $\alpha_{au} = 1$ and $\alpha_{fm} = 10$. The network for auxiliary task G_{au} like the generator and discriminator D are trained in an alternating manner by minimizing the adversarial loss as [40]

$$l_{G_{au}} = -E_{I_{org}}[D(O_{bdr\_pre}, I_{org}] \tag{8}$$

$$\begin{aligned} l_D = &-E_{(I_{bdr\_gt}, I_{org})}[min(0, -1 + D(I_{bdr\_gt}, I_{org}))] \\ &- E_{I_{org}}[min(0, -1 - D(O_{bdr\_pre}, I_{org}))] \end{aligned} \tag{9}$$

The feature-matching loss l_{fm} compares intermediate features that are extracted from layers of the discriminator. This loss stabilizes the training process since it forces the generator G_{au} to generate the stroke boundary map with representations that are similar to the representations of real stroke boundary map. The feature matching loss l_{fm} is defined as

$$l_{fm} = E[\sum_{i=1}^{T} \frac{1}{N_i} ||D^i(I_{bdr\_gt}, I_{org}) - D^i(O_{bdr\_pre}, I_{org})||_1] \tag{10}$$

where N_i indicates the number of elements in i-th layer of the discriminator, and D^i denotes the activation in the i-th layer of the discriminator D. T is the total number of layers.

IV. EXPERIMENTAL RESULTS

A. DATASET AND IMPLEMENTATION DETAILS

For a fair evaluation of previous methods and each DIBCO competition, we utilize nine public data sets for document binarization, including DIBCO [1], [8], [41], [44] and H-DIBCO [42], [45]–[48] images. The total number of these images is 116 images. They are insufficient for supervised deep learning-based methods. Similar to previous papers such as in [4], [6], we utilize external public document binarization datasets such as the Bickley-diary dataset [49], PHIDB [50], and the Synchromedia Multispectral dataset [51] for training. We randomly cropped about 10000 patches of size 288 × 288 from these datasets. 90% of the patches are used as a training set and the remaining patches as a validation set. The training set and validation set are only selected on images independent from the test set. For example, when evaluating H-DIBCO 2018 dataset, the training set and validation set are the datasets without H-DIBCO 2018 dataset.

We employ the Adam optimizer [52] to train our network with its default hyperparameters, where the initial learning rate $lr = 0.001$, $betas = (0.9, 0.999)$, $eps = 1e-8$, $weight\_decay = 0$. The model is trained for 100 epochs with a batch size of 8. The Sigmoid activation function is applied to the output layer, followed by the threshold of 0.5. Our network is implemented based on the publicly available framework Pytorch 1.6.0 and RTX 3090 (with 24GB memory).

B. EVALUATION METRICS

We followed the measures used in DIBO competitions. They adopted four evaluation metrics, F-measure (FM), pseudo-F-measure (Fps) [55], Peak Signal-to-Noise Ratio (PSNR), and Distance Reciprocal Distortion measure (DRD) [56]. In particular, FM and Fps are chosen for evaluation in document image binarization since the distribution of foreground and background classes are often unbalanced. Fps is similar to F-measure but aims to overcome erroneous pixel-based evaluation methods by each relevance pixel weighted based on its distance from stroke boundaries. PSNR measures how close an original image is to its ground-truth image. For example, there is more similarity between the ground-truth image and the predicted image if the PSNR value is high. DRD measures the visual distortion of images. It is correlated with the human visual perception of distortion. A higher value for F-Measure, pseudo-F-Measure, and PSNR indicate a better result, while a lower value for DRD demonstrates better performance.

C. QUANTITATIVE AND QUALITATIVE COMPARISON

To demonstrate the effectiveness and generalization of the proposed model, we choose two benchmark datasets DIBCO 2011 and 2013 that include machine-printed and handwritten document images as well as two recent benchmark datasets DIBCO 2017, and 2018 with complex noise for testing set. Furthermore, these benchmark datasets are also chosen in recent image binarization methods. Thus, we can compare our model to them easily.

We compare our proposed method with other image binarization methods in terms of metrics mentioned in Section IV-B. We can see that our model performs favorably against other methods under all evaluation metrics except the DRD metric in Table 2, Table 4 and the PSNR metric in Table 5. Specifi-

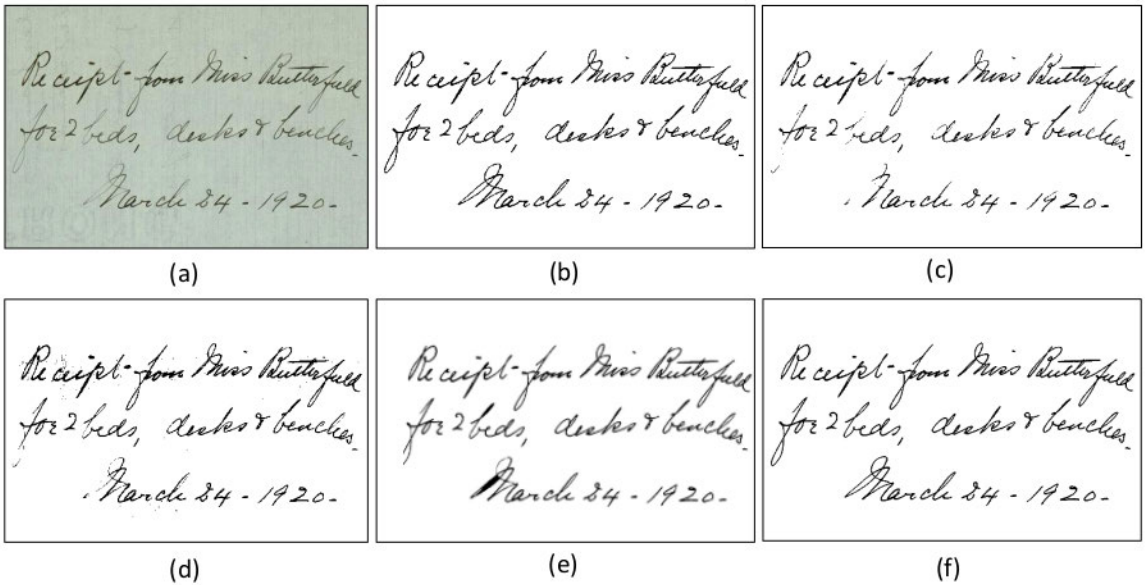

FIGURE 6. Results of binarization methods on the degraded document image with the weak strokes (sample H1 in DIBCO 2011). (a) shows the degraded document image. (b) is ground-truth. (c) is the result of Otsu method [9]. (d) displays the result of Sauvola [13]. (e) is the result of competion Winner [41]. Our proposed method is illustrated in (f).

TABLE 1. Results on DIBCO 2011 dataset.

Methods	FM	Fps	PSNR	DRD
Otsu [9]	82.10	84.79	15.72	8.95
Sauvola [13]	82.10	87.70	15.60	8.50
Competitionwinner [41]	80.9	-	16.1	104.4
GiB [53]	90.33	93.82	18.29	2.99
DSN [6]	93.3	96.4	20.1	2.0
DeepOtsu [4]	93.4	95.8	19.9	1.9
MMN [54]	93.55	96.45	20.10	1.95
cGANs [5]	93.81	95.70	20.26	1.81
CMU-Nets [23]	95.5	-	19.9	1.8
Ours	**95.61**	**97.34**	**22.09**	**1.48**

TABLE 2. Results on DIBCO 2013 dataset.

Methods	FM	Fps	PSNR	DRD
Otsu [9]	83.94	96.52	16.63	10.98
Sauvola [13]	85.02	89.77	16.94	7.58
Competition Winner [44]	92.12	94.19	20.68	3.10
GiB [53]	91.14	94.75	19.58	2.77
DSN [6]	94.4	96.0	21.4	1.8
Younes, et al [57]	94.80	96.65	22.00	1.50
cGANs [5]	95.28	96.47	22.23	**1.39**
MMN [54]	95.43	97.02	22.33	1.46
CMU-Nets [23]	95.88	96.38	22.97	1.46
Ours	**95.96**	**98.13**	**23.14**	1.43

cally, regarding testing on the DIBCO 2011 and 2013 dataset, Table 1 and Table 2 show our proposed method delivers slightly better results than the current state-of-the-art method [23]. Besides, with testing on DIBCO 2017 and 2018, Table 3 and Table 4 illustrate the proposed method has significant performance improvements compared to the traditional methods [9], [13] and recent deep learning methods [5], [6], [23], [44]. Consequently, the proposed method is robust and superior to the existing methods.

TABLE 3. Results on DIBCO 2017 dataset.

Methods	FM	Fps	PSNR	DRD
Otsu [9]	77.73	77.89	13.85	15.54
Sauvola [13]	77.11	84.1	14.25	8.85
Competition Winner [8]	91.04	92.86	18.28	3.40
Suman, et al [58]	83.38	89.43	15.45	6.71
cGANs [5]	90.73	92.58	17.83	3.58
MMN [54]	90.85	93.60	18.50	3.30
CMU-Nets [23]	91.57	93.55	15.85	2.92
Ours	**92.08**	**94.99**	**18.72**	**2.84**

TABLE 4. Results on DIBCO 2018 dataset.

Methods	FM	Fps	PSNR	DRD
Otsu [9]	51.45	53.05	9.74	59.07
Sauvola [13]	67.81	74.08	13.78	17.69
Competition Winner [42]	88.34	90.24	19.11	4.92
Suman, et al [58]	76.84	83.58	15.31	9.58
cGANs [5]	87.73	90.60	18.37	4.58
MMN [54]	89.05	93.65	19.17	4.80
CMU-Nets [23]	89.71	91.62	19.39	**2.51**
Ours	**91.26**	**93.97**	**19.81**	3.42

To prove the ability to preserve strokes in our proposed method, we select samples with disconnected stroke results provided in recent papers for document image binarization methods. The qualitative comparison results are demonstrated in Fig. 6 to 11. According to the figures, the proposed method performs a little better in noise suppression than the existing methods. However, it is essential to note that the proposed method preserves stroke better substantially than the existing methods. In particular, we take failure cases mentioned in [4], [6] about disconnected strokes as Fig. 7(e), 8(e), 9(d) to compare with the corresponding

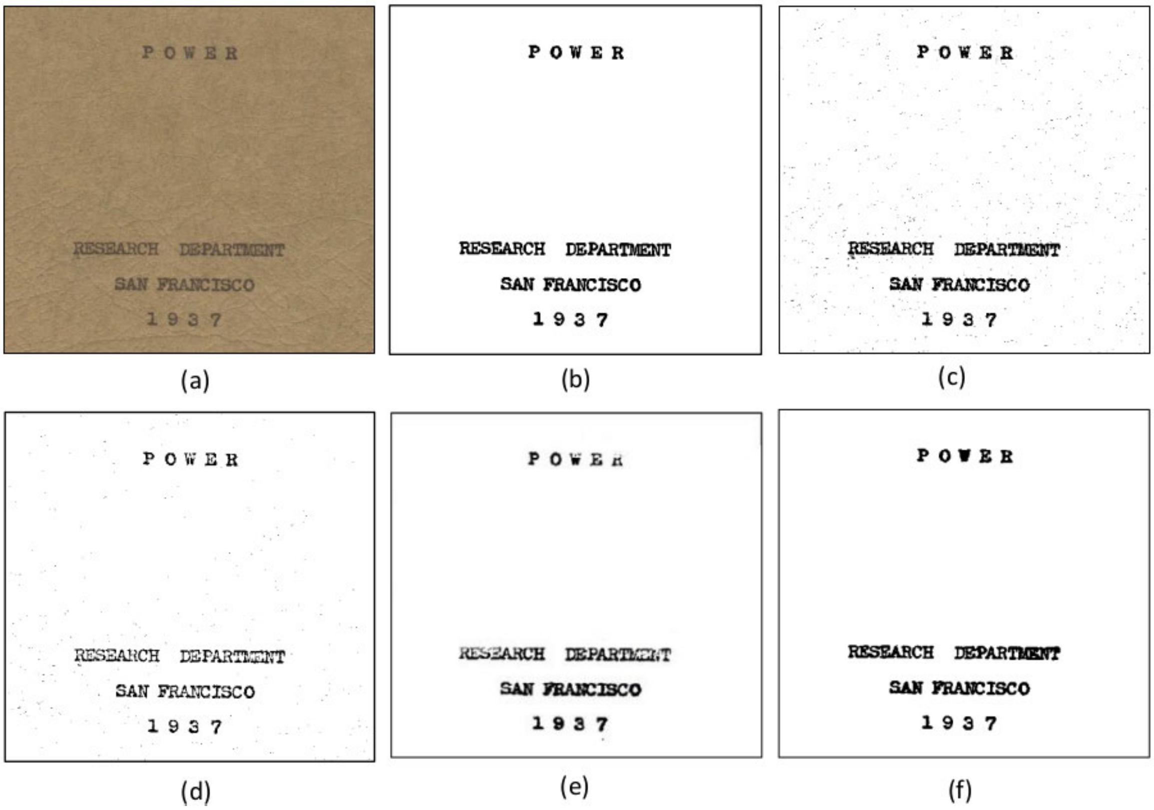

FIGURE 7. Results of binarization methods on the degraded document image with the ambiguous strokes (sample P6 in DIBCO 2011). (a) shows the degraded document image. (b) is ground-truth. (c) is the result of Otsu method [9]. (d) displays the result of Sauvola [13]. (e) is the result of Deepostu [4]. Our proposed method is illustrated in (f).

TABLE 5. Results on DIBCO 2019 dataset.

Methods	FM	Fps	PSNR	DRD
Otsu [9]	52.80	52.56	12.64	24.24
Sauvola [13]	42.52	39.76	7.71	112.40
Competition Winner [43]	72.875	72.15	14.475	16.235
MMN [54]	65.54	64.19	12.95	17.26
Suman, et al [58]	72.87	72.15	**14.48**	16.24
Ours	**73.43**	**73.18**	11.59	**14.64**

results of the proposed method. We observe failure cases about disconnected strokes in [5], [8], [23], [41], [42] as Fig. 6(e), 9(e), 10(c), 11(d), and 11(e), then getting them to compare with the corresponding results of the proposed method. The visual results derived from Ostu [9] and Sauvola [13] are produced by the public Scikit-image library [59]. As a result of the illustrations, we confirm that the proposed method can perform various kinds of noise removal but still preserve strokes better than other image binarization methods.

For the sake of completeness, we additionally choose the most current DIBCO dataset 2019 for testing. It has backgrounds that include more complex noise than other existing DIBCO datasets. As shown in the quantitative result in Table 5 and the qualitative result in Fig. 12, the proposed method achieves superior performance compared to existing methods and better preserves strokes.

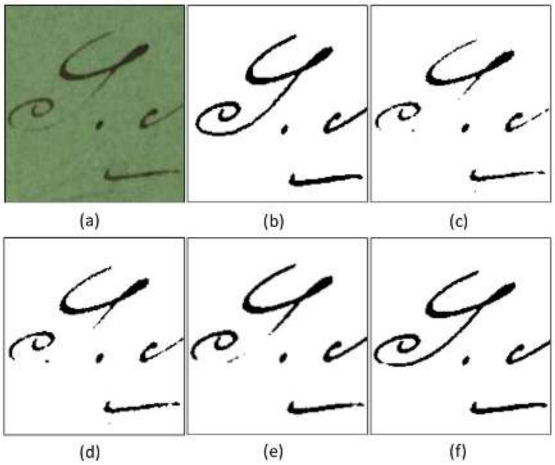

FIGURE 8. Results of binarization methods on the degraded document image with the long and thin strokes (sample H7 in DIBCO 2011). (a) Shows the degraded document image. (b) Is ground-truth. (c) Is the result of Otsu method [9]. (d) Displays the result of Sauvola [13]. (e) Is the result of DSN [6]. Our proposed method is illustrated in (f).

D. ABLATION STUDY

We choose the recent benchmark dataset DIBCO 2017 that includes machine-printed and handwritten document images to prove the effectiveness and generalization in term of preserving stroke of the proposed model.

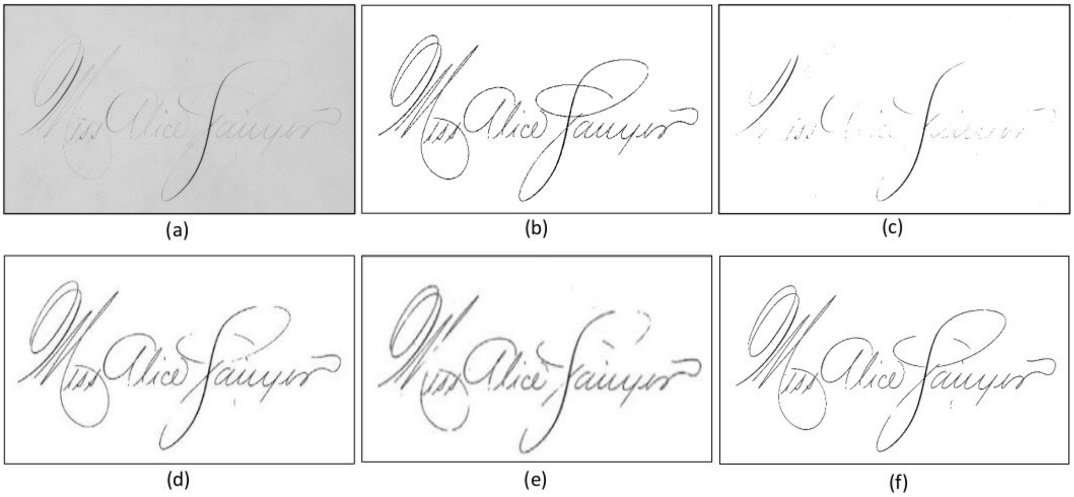

FIGURE 9. Results of binarization methods on the degraded document image with the long and thin strokes (sample HP6 in DIBCO 2013). (a) Shows the degraded document image. (b) Is ground-truth. (c) Is the result of Otsu method [9]. (d) Displays the result of DSN [6]. (e) Is the result of cGANs [5]. Our proposed method is illustrated in (f).

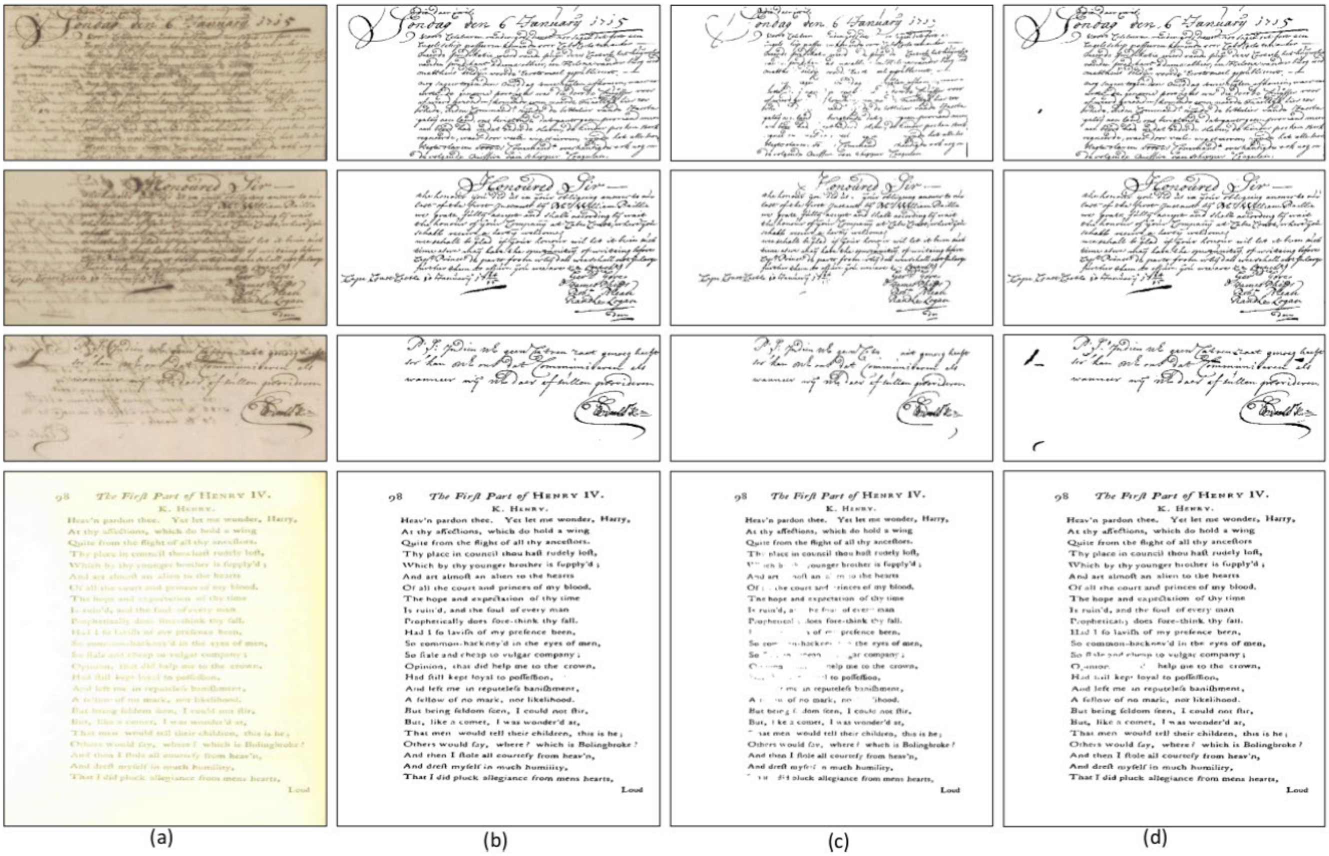

FIGURE 10. Results of binarization methods on the sample images in the DIBCO 2017 dataset. The columns from left to right correspond to the (a) original image, (b) the ground truth, (c) competition Winner [8] and (d) our proposed method.

Baseline's architecture only combines both the shared encoder and the main decoder. The results in Table 6.1 and Fig. 13(c) are acceptable. However, the ambiguous strokes are often disconnected in image binarization results, as shown in the green box.

The architecture from the combination of baseline and stroke boundary feature extracting module, followed by the auxiliary decoder, is multi-task learning. The auxiliary task for stroke boundary prediction implicitly supports the model in preserved stroke and noise removal since the model can

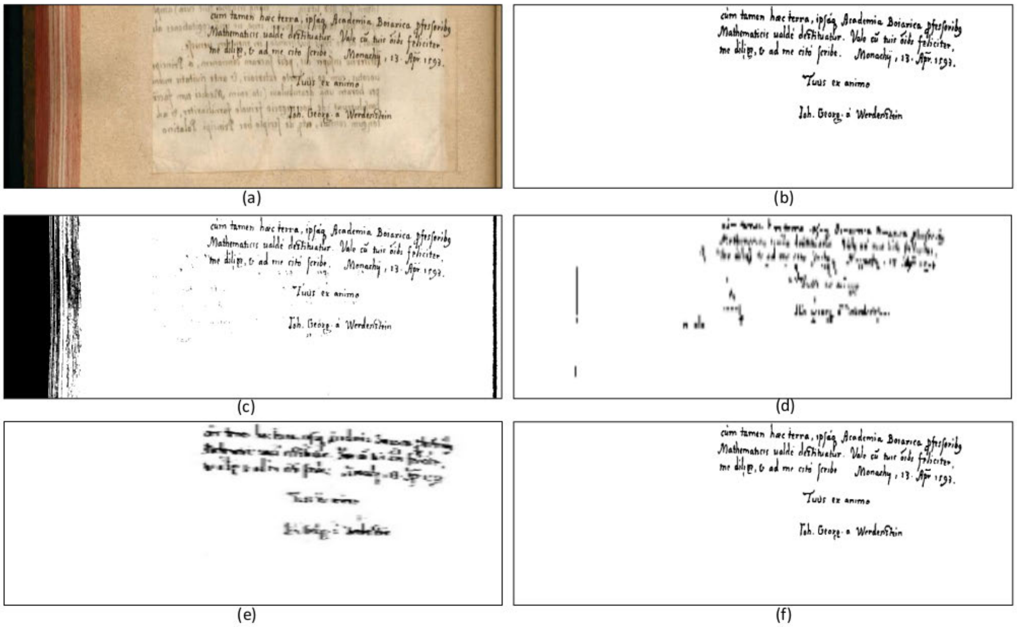

FIGURE 11. Results of binarization methods on the sample image (H3) in the DIBCO 2018 dataset. (a) original image, (b) the ground truth, (c) Ostu method [9], (d) competition Winner [42], (e) CMU-Nets [23] and (f) our proposed method.

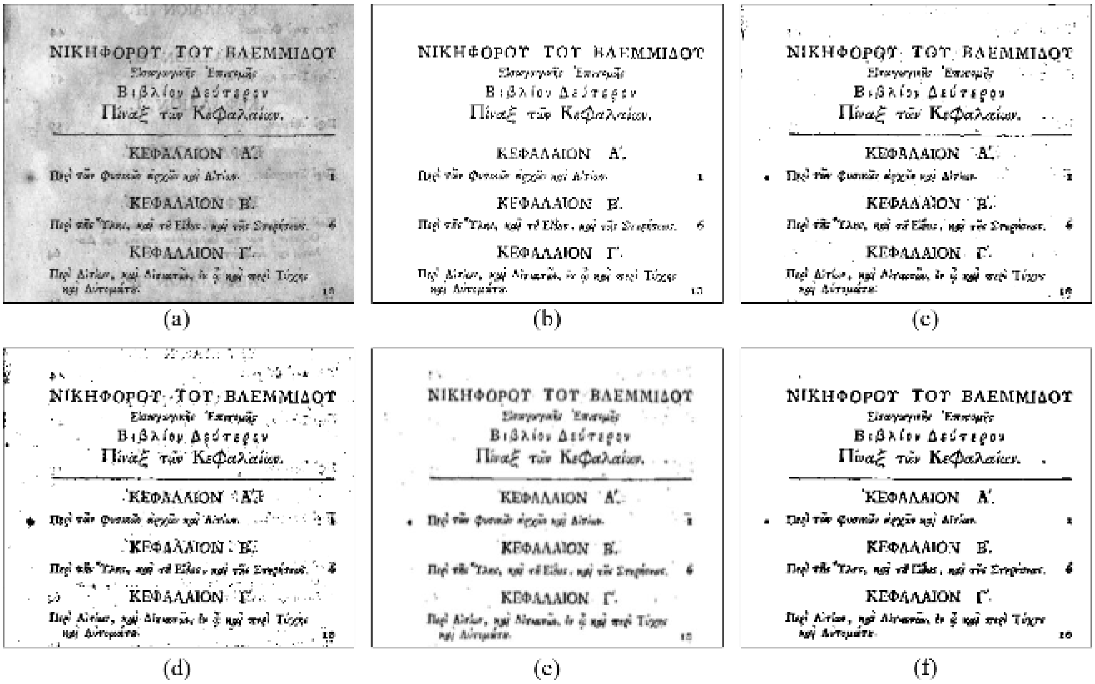

FIGURE 12. Results of binarization methods on the sample image in the DIBCO 2019 dataset. (a) original image, (b) the ground truth, (c) Ostu method [9], (d) Sauvola [13], (e) competition Winner [43], (f) our proposed method.

understand structure information of strokes more, leading to Precision (92.23) and Recall (89.91) increase as shown in Table 6.2 and better-preserved stroke in the green box of Fig. 13(d).

The architecture from the combination of baseline, the stroke boundary feature extracting module followed by the auxiliary decoder, and the feature fusion module can preserve strokes at a higher level thanks to the feature fusion module, as in the green box of Fig. 13(e). This is because the module can suppress the component false negative (FN) in the Recall metric [55], resulting in generating connected strokes. In other words, the Recall metric (91.80) in Table 6.3 is significantly higher than in Table 6.2.

TABLE 6. Ablation study on different models.

Model	Precision	Recall	FM
1. Baseline	91.99	89.66	90.38
2. Baseline+stroke boundary feature extracting module	92.23	89.91	90.81
3. Baseline+stroke boundary feature extracting module+ feature fusion module	91.40	91.80	91.44
4. Baseline+stroke boundary feature extracting module+ feature fusion module+ Discriminator	**92.48**	**91.91**	**92.08**

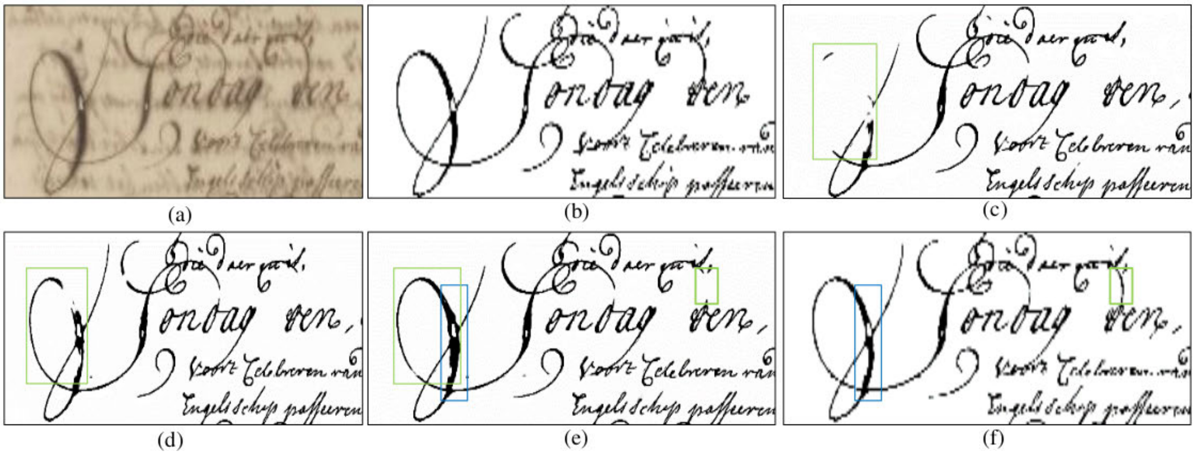

FIGURE 13. Binarization results from the different configurations (a) original image, (b) the ground truth, (c) Baseline, (d) Baseline + stroke boundary feature extracting module, (e) Baseline + stroke boundary feature extracting module + feature fusion module, (f) Baseline + stroke boundary feature extracting module + feature fusion module + discriminator (our proposed method). Green boxes aim to compare in connected stroke level. Blue boxes compare in sharp stroke level.

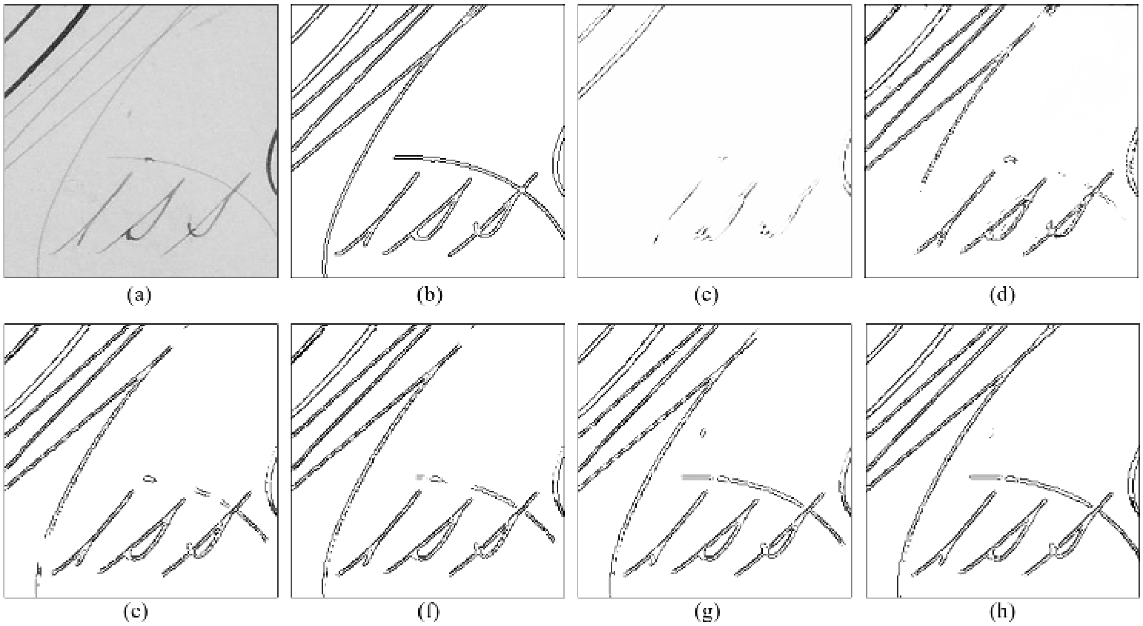

FIGURE 14. Predicted boundary maps of one sample from validation set having weak stroke boundaries at each different epoch during the training phase. (a) The degraded document image. (b) The ground-truth of weak stroke boundary. (c) The result at epoch-1. (d) The result at epoch-20. (e) The result at epoch-40. (f) The result at epoch-60. (g) The result at epoch-80. (h) The result at epoch-100.

Table 6.4 and Fig. 13(f) are the results of our proposed method. In particular, the blue box in Fig. 13(e) illustrates connected strokes. However, many redundant parts locate on stroke, leading to rough connected strokes. In the blue box in Fig. 13(f), it shows the connected strokes sharply as the result of the Precision metric increase in Table 6.3 compared with Table 6.2 because the proposed model suppresses the component false positive (FP) in Precision metric [55].

Besides, the green box in Fig. 13(f) shows preserving strokes better than Fig. 13(e). It achieves preserved stroke sharply because the proposed method employs the stroke boundary information-based discrimination network more for adversarial training. The Precision metric increases moderately by 0.49 from 91.99 of the baseline model to 92.48 of the proposed model, while the Recall metric has a significant increase of 2.25 from 89.66 of the baseline model to 91.91 of the proposed model. This result proves that the proposed method has an outstanding ability to preserve stroke in document image binarization. Finally, the overall metric FM (92.08) of the proposed method is higher than other models.

To further demonstrate the contribution of the proposed method, we can visualize output of the stroke boundary feature extracting module during the training phase. Since the stroke boundary features have too many channels (the number of channels = 64) in every different epoch, we simplify the feature representation into a single channel by feeding the features into the auxiliary decoder to produce the corresponding feature map. The output map has a single channel and the same resolution as the input image. Therefore, the map can be visually compared to the ground truth easily in terms of preserved boundaries. The visualizing result is derived from the input image of the validation set having weak stroke boundaries which are usually thin and long, thus difficult to correctly binarize. As in Fig. 14, the predicted maps of weak stroke boundaries show the result get closer to the ground truth in Fig. 14 (b) gradually through increasing epochs.

V. CONCLUSION

This paper presents a novel document image binarization model with weak or ambiguous strokes that are often disconnected after the binarization process in existing methods. To preserve the strokes in degraded document images after the binarization process, we embed structure information of strokes into the binarization network. Based on this idea, we propose an auxiliary task to learn structure information in terms of stroke boundary features in an adversarial manner and integrate the learned boundary features into the main task for document image binarization. Firstly, the auxiliary task utilizes additional local edge features and shared global location features to obtain stroke boundary features. Secondly, we leverage boundary ground truth to supervise the obtained stroke edge feature in the auxiliary task in an adversarial manner. The adversarial training method is to embed expert knowledge about structure information in the model. Thirdly, the learned boundary feature from the auxiliary task supports the main task directly. The feature fusion module refines the main task again to produce the final image binarization results. Experiments demonstrate that our method achieves better-preserved stroke and better performance than existing methods on benchmark H-DIBCO and DIBCO datasets. In the future, we expect to extend the potential of the proposed model to be used for semantic segmentation tasks with ambiguous objects that are often vanished after the segmentation process.

VI. DECLARATION OF COMPETING INTEREST

The authors confirm that all authors of this manuscript have no conflicts of interests to declare.

REFERENCES

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

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

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[4] S. He and L. Schomaker, "DeepOtsu: Document enhancement and binarization using iterative deep learning," *Pattern Recognit.*, vol. 91, pp. 379–390, Jul. 2019.

[5] J. Zhao, C. Shi, F. Jia, Y. Wang, and B. Xiao, "Document image binarization with cascaded generators of conditional generative adversarial networks," *Pattern Recognit.*, vol. 96, Dec. 2019, Art. no. 106968.

[6] 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.

[7] R. Caruana, "Multitask learning," *Mach. Learn.*, vol. 28, no. 1, pp. 41–75, 1997.

[8] 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)*, Nov. 2017, pp. 1395–1403.

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

[10] J. Kittler and J. Illingworth, "Minimum error thresholding," *Pattern Recognit.*, vol. 19, no. 1, pp. 41–47, Jan. 1986.

[11] A. D. Brink, "Thresholding of digital images using two-dimensional entropies," *Pattern Recognit.*, vol. 25, no. 8, pp. 803–808, Aug. 1992.

[12] W. Niblack, "An introduction to digital image processing, 215 strandberg publishing company," *Copenhagen, Denmark*, 1985.

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

[14] 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.

[15] 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.

[16] X. Zhang, C. He, and J. Guo, "Selective diffusion involving reaction for binarization of bleed-through document images," *Appl. Math. Model.*, vol. 81, pp. 844–854, May 2020.

[17] Z. X.-S. C. Shu-Zhen, "Image segmentation based on global binarization and edge detection," *J. Comput. Aided Design Comput. Graph.*, vol. 2, pp. 118–121, Feb. 2001.

[18] Q. Chen, Q.-S. Sun, P. Ann Heng, and D.-S. Xia, "A double-threshold image binarization method based on edge detector," *Pattern Recognit.*, vol. 41, no. 4, pp. 1254–1267, Apr. 2008.

[19] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

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

[21] C. Tensmeyer and T. Martinez, "Document image binarization with fully convolutional neural networks," in *Proc. 14th IAPR Int. Conf. Document Anal. Recognit. (ICDAR)*, vol. 1, 2017, pp. 99–104.

[22] 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.

[23] S. Kang, B. K. Iwana, and S. Uchida, "Complex image processing with less data—Document image binarization by integrating multiple pre-trained U-Net modules," *Pattern Recognit.*, vol. 109, Jan. 2021, Art. no. 107577.

[24] X. Qin, Z. Zhang, C. Huang, C. Gao, M. Dehghan, and M. Jagersand, "BASNet: Boundary-aware salient object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 7479–7489.

[25] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*. [Online]. Available: http://arxiv.org/abs/1511.07122

[26] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: http://arxiv.org/abs/1502.03167

[27] R. H. R. Hahnloser, H. S. Seung, and J.-J. Slotine, "Permitted and forbidden sets in symmetric threshold-linear networks," *Neural Comput.*, vol. 15, no. 3, pp. 621–638, Mar. 2003.

[28] P. Zhang, D. Wang, H. Lu, H. Wang, and X. Ruan, "Amulet: Aggregating multi-level convolutional features for salient object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 202–211.

[29] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1395–1403.

[30] 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

[31] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 818–833.

[32] A. Mahendran and A. Vedaldi, "Understanding deep image representations by inverting them," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 5188–5196.

[33] J. Zhao, J.-J. Liu, D.-P. Fan, Y. Cao, J. Yang, and M.-M. Cheng, "EGNet: Edge guidance network for salient object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2019, pp. 8779–8788.

[34] H. J. Lee, J. U. Kim, S. Lee, H. G. Kim, and Y. M. Ro, "Structure boundary preserving segmentation for medical image with ambiguous boundary," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 4817–4826.

[35] P.-T. de Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Ann. Operations Res.*, vol. 134, no. 1, pp. 19–67, Feb. 2005.

[36] G. Mattyus, W. Luo, and R. Urtasun, "DeepRoadMapper: Extracting road topology from aerial images," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 3438–3446.

[37] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. Thrity-Seventh Asilomar Conf. Signals, Syst. Comput.*, 2003, pp. 1398–1402.

[38] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," 2018, *arXiv:1802.05957*. [Online]. Available: http://arxiv.org/abs/1802.05957

[39] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional GANs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8798–8807.

[40] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 7354–7363.

[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, 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)*, Aug. 2018, pp. 489–493.

[43] 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)*, Sep. 2019, pp. 1547–1556.

[44] 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.

[45] 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.

[46] 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.

[47] 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.

[48] 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.

[49] 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.

[50] 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.

[51] 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. ICDAR*, Aug. 2015, pp. 1181–1185.

[52] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[53] S. Bhowmik, R. Sarkar, B. Das, and D. Doermann, "GiB: A game theory inspired binarization technique for degraded document images," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1443–1455, Mar. 2019.

[54] Y. Akbari, S. Al-Maadeed, and K. Adam, "Binarization of degraded document images using convolutional neural networks and wavelet-based multichannel images," *IEEE Access*, vol. 8, pp. 153517–153534, 2020.

[55] 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.

[56] H. Lu, A. C. Kot, and Y. Q. Shi, "Distance-reciprocal distortion measure for binary document images," *IEEE Signal Process. Lett.*, vol. 11, no. 2, pp. 228–231, Feb. 2004.

[57] 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.

[58] S. K. Bera, S. Ghosh, S. Bhowmik, R. Sarkar, and M. Nasipuri, "A non-parametric binarization method based on ensemble of clustering algorithms," *Multimedia Tools Appl.*, vol. 80, pp. 7653–7673, Oct. 2021.

[59] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, and V. Dubourg, "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Nov. 2011.

QUANG-VINH DANG received the B.E. degree in mechatronics technology from the Ho Chi Minh University of Technical Education, in 2011, and the master's degree in mechatronics engineering from the Ho Chi Minh City University of Technology, in 2013. He is currently pursuing the Ph.D. degree in artificial intelligence convergence with Chonnam National University, South Korea. His research interests include computer vision and deep learning.

GUEE-SANG LEE (Member, IEEE) received the B.S. degree in electrical engineering and the M.S. degree in computer engineering from Seoul National University, South Korea, in 1980 and 1982, respectively, and the Ph.D. degree in computer science from Pennsylvania State University, in 1991. He is currently a Professor with the Department of Artificial Intelligence Convergence, Chonnam National University, South Korea. His research interests include image processing, computer vision, and video technology.