Learning 2D Morphological Network for Old Document Image Binarization

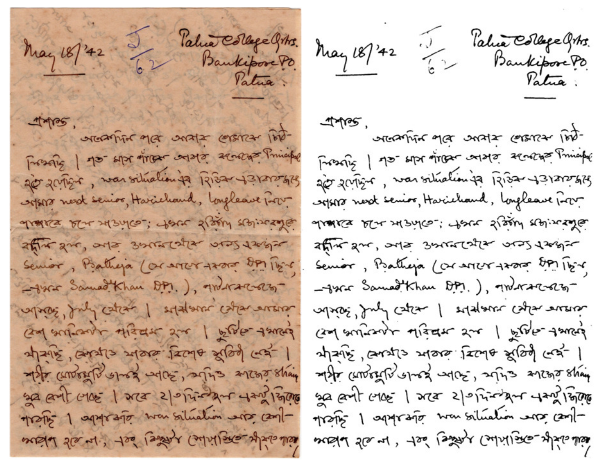
Ranjan Mondal
Indian Statistical Institute, Kolkata
ranjan15_r@isical.ac.in

Deepayan Chakraborty
Indian Statistical Institute, Kolkata
deepayan504@gmail.com

Bhabatosh Chanda
Indian Statistical Institute, Kolkata
chanda@isical.ac.in

***Abstract*—Document image binarization, especially old handwritten documents, is a very important yet challenging task. There are various bottlenecks for binarizing historical documents due to different types of degradations present simultaneously such as back impression, ink bleed through, faded colours, and wear and tear of the writing media. We consider these degradations as various types of noise in the document image. Here we have proposed a 2D morphological network which consists of basic morphological operation like dilation and erosion to perform our targeted task. The network also includes linear combination of output from dilation and erosion operations. The aforementioned 2D morphological network is applied for image binarization, where the structuring elements (SEs) and the weights of the linear combination layer are learned through back-propagation. The proposed network has been evaluated on DIBCO 2017 and H-DIBCO 2018 and ISI-Letter dataset. Our results show more convincing as compared to the results of other state-of-the-art methods. Though the network is developed for old handwritten documents, it may be tuned to work for old printed documents also. The source code can be found here https://github.com/ranjanZ/ICDAR_Binarization**

***Keywords*-Morphological Networks, Morphological Neurons, Document Image Processing.**



(a) Input Image (b) Our Output

Figure 1: Input image and the output from our proposed 2D Morphological Network

I. Introduction

The world of twenty-first century is entirely shifting towards a paperless world because of today's advancement in technology. Although the world is moving towards paperless, the importance of the historical documents remained the same. Historical documents may be a family letter, a document of national importance or old machine printed valuable book. All these types of documents describe the socioeconomic condition of that period. Besides those information, the handwritten documents may also describe the personality and mood of the writer. For extracting the information from the historical documents, binarization is an important preprocessing step. But the aging of the pages, bleeding of ink, wear and tear of the writing medium and fading of the inks cause deterioration and noise in the binarized image.

Binarization and elimination of noise in the historical documents is a challenging task, because of their varying degradation types. Different types of degradation add different types of noise in the binarized image. The task becomes even more difficult when the intensity of the ink colour becomes closer to the intensity of the background colour. Researchers have started working on document image binarization decades ago, and there are various classical binarization methods available which compute local and global threshold for binarization [8], [13]. But efficient and useful automatic methods for binarizing old (handwrtten) documents are yet to be developed. Recently deep convolutional network is being widely used for binarization and noise removal.

Mathematical morphological methods are very popular to filter out different structures or noise in an image. Morphological methods produce remarkable results in image processing in which the type filter(s) as well as the shape and size of structuring elements (SEs) or kernels are designed properly. Recently Dense morphological network has been proposed by Mondal *et. al* [7] which can learn 1D structing elements. It has a layer with morphological neurons followed by a linear combination layer. They have also shown the network which this structure can approximate any continuous function. However, the network can only process one dimensional feature vector. With the motivation to process 2D image data with morphological network we have proposed 2D Morphological Network.

In this paper our contribution may be summarized as:
1. We have defined 2D morphological Network for image processing tasks and utilize this network for binarization of

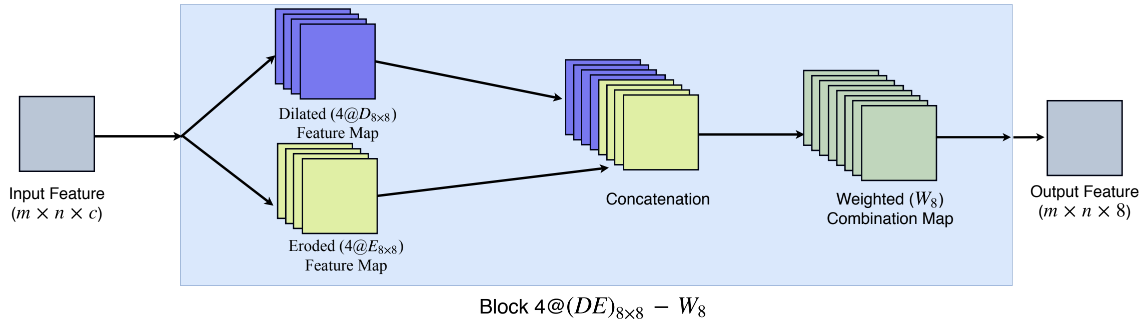


Figure 2: Architecture of single block of 2D morphological layer. It contains a parallel dilation and erosion layer, followed by Concatenation layer and a linear combination layer at the end. The dilation-erosion layer 4 dilated and eroded feature map is constructed by 4 different dilation and erosion operation on input feature with the structuring elements (SE) of size 8×8. In the next layer 8 different linear combination is done and that also serves as the output feature map. We denote this block as $4@(DE)_{8\times8} - W_8$ which says parallel 4 dilation and 4 erosion of size 8×8 is done followed by 8 different linear combination.

old (handwritten) documents [see Figure 1]. SEs are learned using by back propagation method by minimizing a simple loss function.
2. We have made use of universal approximation power of Dense Morphological Network and proposed a morphological block consisting of dilation and erosion layers, and also a linear combination layer.
3. The proposed 2D Morphological Network is built by concatenating these morphological blocks, which are capable to learn the spatial features from the input image.

II. Related Works

Binarization of document image is a classical problem. The main aim of binarization is to convert each pixel either foreground or background properly. But addition of noise leads to incorrect binarization. It is very obvious that the process becomes more challenging for a historical document which contains various types of noises. Various approaches has been proposed by the researchers to binarize and de-noise historical document images. Dubois and Pathak [3] have proposed a model which is capable of reducing the bleed through noise from gray scaled document images. This model works on both side scanned grayscaled images. The images are flipped and transformed in order to find the right alignment of the foreground of the both side of the page. Moghaddam and Cheriet [6] have proposed a restoration method for single-sided low-quality document images (RSLDI). This method has enhanced the quality of diffusion-based restoration method by introducing multi-lavel classifier. This method is capable of removing the bleed through or similar noises from single-sided grayscaled document images and directly accessing the information from them. Rivest-Hnault *et. al.* [12] have proposed a local linear level set method for binarization and restoration of degraded document images. The level set methodology integrates the document related aspects, is combined with local linear model to detect the configurations of the smoothly varying intensities of the ink strokes. This model is capable of binarizing the historical documents after de-noising them. Jia *et. al.* [4] have proposed a model of restoration and binarization of historical documents by analyzing the structural symmetry of the strokes. This model includes a background removal method followed by calculation of an adoptive threshold and extraction of structural symmetric pixels (SSP). The aforementioned methods have a common drawback. The methods are capable of handling some particular types of noises, instead of handling all types of noises in a single model. In this work basic concepts of morphological network. Morphological perceptron was first introduced by Davidson *et. al.* [2] for the template identification problem. Later, Ritter*et. al.* [11] tried to solve classification problem by a single layer architecture with morphological neurons where the learned decision boundaries are parallel to the axes. Recently, Mondal *et. al.* [7] proposed a dense morphological network (DenMo-Net) with the linear combination of morphological neurons. But the netowk is limited to only 1D feature vector as input. Using the properties of DenMo-Net we have defined our 2D morphological Network. In the next section we have describied the basic building block of our 2D Morphological Network. In the later section we have used the building block to build a network and applyed it on document image binarization problem.

Rest of the paper is arranged as follows. In section III, we have described our proposed morphological network in details, section IV describes how we have carried out the experiment and at last in section V we have concluded the paper.

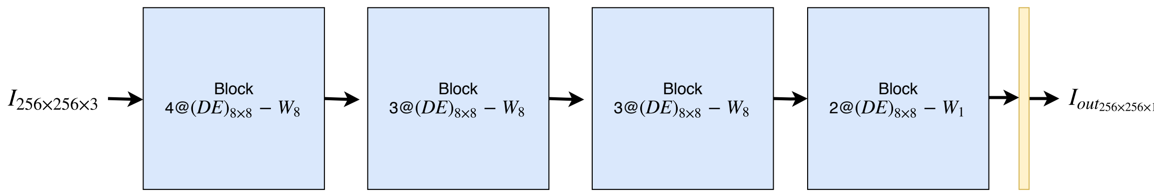


Figure 3: Network with series of four 2D morphological block followed by a sigmoid activation function. In this network input I is taken as image patch of shape (256,256,3) where as output is I_{out} is of shape (256,256,1). Each pixel of I_{out} denotes probability that the pixel if I belongs to text.

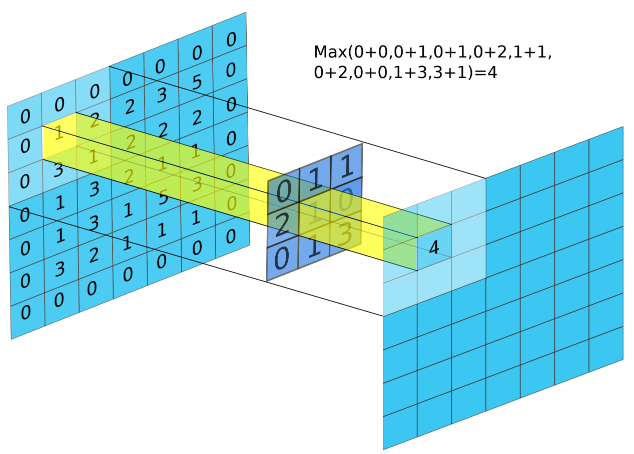


Figure 4: Dilation operation on input of size (7,7,1) input with SE size (3,3)

III. Proposed Method

Recently Dense Morphological Network [7] consist of a layer of dilation and erosion neurons followed by a linear combination layer. It is shown that a single dilatation-erosion layer with enough dilation erosion neurons followed by a linear combination layer can approximate any continuous function. However, processing image with that network is very expensive due to huge image dimension. Network is also unaware of spatial structure in the image data. With this motivation, we have extended the idea of Dense Morphological Network and have defined a 2D morphological network for the image processing task.

In the following section we have defined a basic morphological block (or module) which can process a 2D image feature map. In this work we have concatenated a series of such morphological blocks and have built a network for image binarization task. Colour of pixel is used as the feature I here. However, other features may also be used.

A. 2D Morphological block

Let I denotes input feature map of size $(m \times n \times c)$. Let $W_d \in R^{a \times b \times c}$ and $W_e \in R^{a \times b \times c}$ are dilation and erosion kernels or SEs respectively of size $a \times b \times c$ that are applied I. It may be noted that $a \leq m$ and $b \leq n$. Dilation $(\oplus)$ and erosion $(\ominus)$ operation on image I is defined as the following

$$(I \oplus W_d)(x, y) = \max_{i \in S_1 j \in S_2 k \in S_3} (I(x+i, y+j, k) + W_d(i, j, k)) \quad (1)$$

$$(I \ominus W_e)(x, y) = \min_{i \in S_1 j \in S_2 k \in S_3} (I(x+i, y+j, k) - W_e(i, j, k)) \quad (2)$$

Where $S_1 = \{1, 2, .., a\}$, $S_2 = \{1, 2, .., b\}$ and $S_3 = \{1, 2, .., c\}$. In figure 4 we have graphically shown the working of dilation operation. If c_1 number of dilation and c_2 erosion kernels are used then the output feature map I_1 has size $(m \times n \times c_3)$, where $c_3 = c_1 + c_2$. Note that in our task we have always used zero padding.

In the next layer, the feature maps of I_1 are combined linearly as

$$I_2(x, y) = b + \Sigma_{k=1}^{c_3} w_k I_1(x, y, k) \quad (3)$$

where w_k are the weights of the linear combination and b is bias. If λ be the number of such linear combinations, output size would be $(m \times n \times \lambda)$. We call dilation-erosion layer followed by linear combination layer together a *2D Morphological module* or *2D Morphological block*. Figure 2 shows such a block with $a = b = 8$, $c_1 = c_2 = 4$ and $\lambda = 8$.

The motivation behind taking this two layers (dilation-erosion and linear combination) consecutively, because it exactly matches basic principle of dense morphological network. On the other hand, applying dilation and erosion on image reduces dark and bright noise respectively. Their linear combination results in removal of both kinds of noise.

In the next section we show how series of 2D morphological blocks can binarize old document images.

B. Morphological Network for Image Binarization

Each block of this morphological network does some kind of spatial image processing and generates new feature map. Concatenation of multiple such morphological modules with proper training may generate desired feature map in spatial domain. As shown in figure 3 we have taken four such blocks in series, where input image patch I is of size $(256\times256\times3)$ and output I_{out} is of size $(256 \times 256 \times 1)$. Each pixel of I_{out} denotes the probability that the pixel belong to text. Note that in the last layer we have used sigmoid activation function to bound the output in [0,1].

C. Training of Morphological Network

All the parameters of our 2D morphological network, like SEs in the dilation-erosion layer and weights of the linear combination layer are randomly initialized. We learn those parameter in a supervised way. For the document image binarization we have used mean squared error as a loss. In the dilation-erosion layer dilation and erosion operations are defined by max and min respectively. All though max or min are not fully differentiable, it is piecewise differentiable. So, we can compute piecewise derivative of the loss with respect to the SEs. Hence, through back-propagation algorithm all the SEs and the weights of the linear combination are learned.

IV. Experiments and Results

The proposed 2D morphological network is implemented[1] in python by using library Keras with TensorFlow backend. The experiment is carried out on a machine with Intel Xeon 16 core processor and 128GB RAM. We have evaluated our proposed 2D morphological Network for document image binarization on publicly available dataset, namely DIBCO 2017 [9], H-DIBCO 2018 [10] and also on ISI-letters dataset [5]. DIBCO 2017 and H-DIBCO 2018 datasets contain 20 and 10 historical document images respectively. ISI-letter dataset contains 26 letters from the time of second world war that are of different sizes varying from 1499×2343 to 2255×2821. Among the 26 images we have adopted 10 images for our experiment. For the training of the 2D Morphological Network we have used the samples from DIBCO 2009 and H-DIBCO 2016 dataset and also 50% samples of ISI-letters dataset.

For data preparation, high resolution document images are split into over lapping patches of size 256×256 with stride 50. The patches are then individually fed into the network for prediction of binarized images. The output image is reconstructed by mosaicing the patches properly. The overlapped areas are averaged to implement superposition.

The network optimization settings and evaluation matrices are described. Then we present qualitative and quantitative evaluation on test data.

A. Optimization of 2D Morphological Network and Evaluation Metrices

All the trainable parameters in each layer of 2D Morphological Network are initialized by standard glorot uniform initializer. In this work we have used mean squared error between the predicted output and the ground truth as loss. For the minimization of loss, Adam optimizer with default parameter settings ($lr = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1e{-8}$) is employed. Back-propagarion algorithm is used to update the SEs and the weights in the linear combination layers. The network is trained up to 100 and 20 epochs on DIBCO and ISI-Letter dataset with batch size of 4.

For the evaluation of our network, we have used F-Measure (FM), Pseudo F-Measure(Fps), Peak Signal to Noise Ratio (PSNR), DRD and Structural Similarity Index (SSIM) [14].

[1]https://github.com/ranjanZ/ICDAR_Binarization

Table I: Qualitative measure of binarization of our proposed approach on DIBCO 2017 dataset.

Image	FM	Fps	PSNR	DRD	SSIM
1	83.87	87.24	15.65	4.26	0.92
2	88.34	92.64	16.55	4.16	0.92
3	89.08	93.78	18.02	3.13	0.95
4	85.12	89.07	17.75	4.68	0.95
5	86.19	90.69	20.11	4.18	0.97
6	93.12	95.81	15.19	2.74	0.87
7	92.47	94.94	15.08	3.08	0.87
8	91.82	95.12	19.01	2.72	0.93
9	91.81	91.90	17.55	2.51	0.93
10	94.01	94.61	17.57	2.31	0.92
11	94.44	94.80	18.08	3.63	0.92
12	88.50	90.31	16.76	4.10	0.92
13	83.70	83.64	15.89	8.26	0.88
14	90.72	92.02	18.43	3.38	0.94
15	93.20	93.94	16.80	2.48	0.91
16	95.23	95.96	22.78	1.96	0.98
17	79.93	82.00	17.37	4.90	0.94
18	86.44	92.08	15.92	4.82	0.90
19	91.10	93.94	18.82	3.08	0.95
20	88.97	91.16	15.74	4.69	0.88
Average	**89.40**	**91.78**	**17.45**	**3.75**	**0.92**

Table II: Qualitative measure of binarization of our proposed approach on H-DIBCO 2018 dataset.

Image	FM	Fps	PSNR	DRD	SSIM
1	78.93	81.58	16.08	12.04	0.94
2	78.70	85.30	17.20	7.09	0.94
3	90.49	94.71	15.24	4.08	0.90
4	74.06	78.65	17.32	8.50	0.90
5	70.76	74.10	14.01	16.69	0.88
6	95.72	96.21	20.79	2.63	0.97
7	85.57	88.31	19.92	4.61	0.97
8	90.16	90.64	16.42	3.33	0.93
9	87.95	87.83	16.89	7.83	0.94
10	75.17	81.32	11.92	12.09	0.84
Average	**82.75**	**85.90**	**16.58**	**7.89**	**0.92**

B. Quantitative Results

We have evaluated the outputs of our proposed approach using the aforementioned evaluation metrics (FM, Fps, DRD, SSIM, PSNR). As stated earlier, we have used three datasets (DIBCO 2017, H-DIBCO 2018 and ISI-Letter) and tabulated the metric values along with their averages in

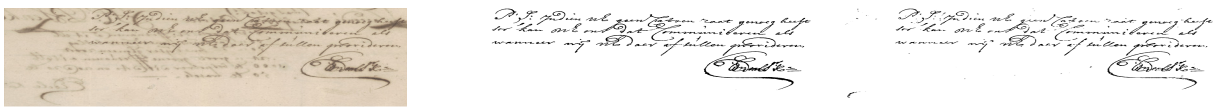


Figure 5: Sample outputs of our proposed binarization approach using handwritten and machine printed documents of DIBCO 2017 dataset.

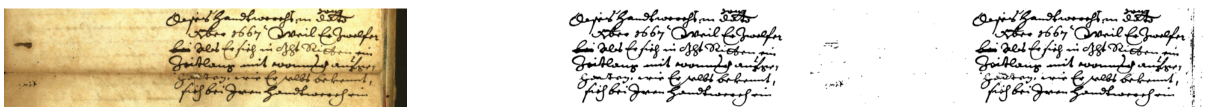


Figure 6: Sample outputs from our proposed 2D morphological Network on H-DIBCO 2018 dataset.

Table III: Qualitative measure of binarization of our proposed approach on ISI-Letter dataset.

Image	FM	SSIM	PSNR
1	98.24	0.94	17.71
2	99.05	0.96	20.27
3	87.64	0.71	9.37
4	98.47	0.95	18.41
Average	**95.85**	**0.89**	**16.44**

Table I, Table II and Table III respectively. The average FM, PSNR, Fps and DRD values of our proposed method is quantitatively compared with the top 6 approaches reported in DIBCO 2017 and H-DIBCO 2018 competitions of document image binarization. Our method is also quantitatively compared with the results achieved by applying the approach by Quang *et. al.*[1]. We have carried out the comparison based on the measures, FM, Fps, PSNR, DRD and tabulated them in Table IV and Table V. Table IV reveals that for DIBCO 2017 dataset our method achieves satisfactory metric values. For H-DIBCO 2018 (Table V) dataset, the FM, PSNR, Fps values achieved by our method are better than that of the first runner up. Moreover the FM and PSNR values achieved are better than that of Quang *et. al.*. As we are comparing on binary images, F-measure (FM) is more indicative for the accuracy than PSNR. From this perspective, our method is found to achieve overall better accuracy than other recently reported methods.

Even for the ISI-letter dataset our method achieves a convincing quantitative measures (see Table III).

C. Qualitative Results

We have also provided some sample outputs along with their ground truth images in Figure 5 and Figure 6 for DIBCO 2017 and H-DIBCO 2018 datasets. From the figures it is clear that our approach is able to binarize the handwritten as well as machine printed documents even if various types of degradations including back impression and page

Table IV: Comparison of the quantitative measurements of our proposed approach with top 6 methods reported in ICDAR2017 Competition on Document Image Binarization on DIBCO 2017 dataset[9].

Rank	Method	FM	Fps	PSNR	DRD
1	Ilin *et. al.*	91.04	92.86	18.28	3.40
2	Zhang *et. al.*(Method a)	89.67	91.03	17.58	4.35
3	Vo *et. al.*	89.42	91.52	17.61	3.56
4	Tensmeyer (Method a)	86.05	90.25	17.53	4.52
5	Tensmeyer (Method b)	83.76	90.35	17.07	4.33
6	Zhang *et. al.*(Method b)	88.37	89.59	17.10	4.94
-	Quang *et. al.*	89.71	92.57	17.71	3.45
-	**2DMorphological Network**	89.40	91.78	17.45	3.75

Table V: Comparison of the quantitative measurements of our proposed approach with top 6 methods reported in ICFHR2018 Competition on Document Image Binarization on H-DIBCO 2018 dataset [10].

Rank	Method	FM	Fps	PSNR	DRD
1	Wei *et. al.*	88.34	90.24	19.11	4.92
2	Adak *et. al.*	73.45	75.94	14.62	26.24
3	Syed Ahsen Raza	70.01	74.68	13.58	17.45
4	Gattal and Djeddi (Method b)	64.52	68.29	13.57	16.67
5	Saddami *et. al.*	46.35	51.39	11.79	24.56
6	adie and Mousa	56.08	60.68	11.5	28.99
-	Quang *et. al.*	82.24	86.03	16.28	8.87
-	**2DMorphological Network**	82.75	85.90	16.58	7.89

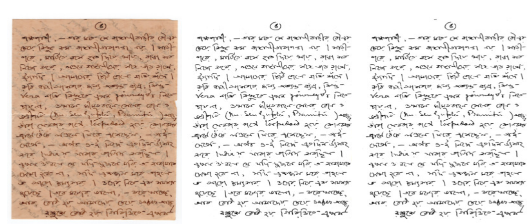
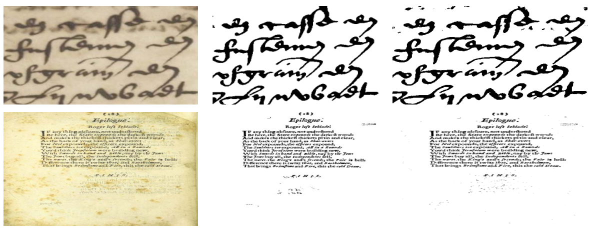
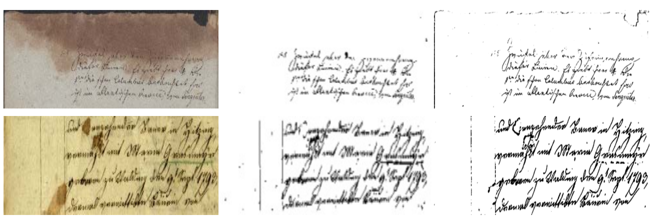


Figure 7: Sample outputs of our proposed binarization approach using ISI-Letters dataset.



Original Image Winners Output Ours Output

Figure 8: Comparison of the outputs of our proposed approach with that of the winner of the ICDAR2017 competition on document image binarization using DIBCO2017 dataset.



Original Image Winners Output Ours output

Figure 9: Comparison of the outputs of our proposed method with the winner of the ICFHR2018 competition on document image binarization using H-DIBCO2018 dataset.

folding marks are present. Moreover, the approach is able to extract the pen strokes properly even if the difference between the foreground and background colour intensity is very low (Figure 5, Figure 6). We have also provided the outputs of our approach on handwritten letters from ISI archive in Figure 7.

We have compared the output images visually with the output of the winning method for DIBCO 2017 and H-DIBCO 2018 datasets (Figure 8, Figure 9). For the outputs of DIBCO 2017 handwritten and machine printed documents (Figure 8), the output of our proposed approach is almost equally clear as the output of the winner approach. Whereas for H-DIBCO 2018 (Figure 9) the figure clearly depicts that the output of our approach is equally good with the winner approach and visually is more noise free than the output of the approach in the second position.

V. Conclusion

In this work we have proposed a 2D Morphological Network and applied it for binarization of old document images. We have shown mostly better results compared to the other methods. Although, in this work, 2D morphological network is exploited for the binarization of document images, it can be applied for other image processing tasks too. Here we have implemented the proposed 2D Morphological Network using Dilation and Erosion operators. However, we believe that similar 2D Morphological Network can be constructed using other type of morphological neighbourhood operators, such as opening-closing and hit-and-miss. We consider this as future scope of work.

References

[1] *Binarization of degraded document images based on hierarchical deep supervised network*, 2017.

[2] Jennifer L. Davidson and Frank Hummer. Morphology neural networks: An introduction with applications. *Circuits, Systems and Signal Processing*, 12(2):177–210, June 1993.

[3] Eric Dubois and Anita Pathak. Reduction of bleed-through in scanned manuscript documents. In *PICS*, volume 1, pages 177–180, 2001.

[4] Fuxi Jia, Cunzhao Shi, Kun He, Chunheng Wang, and Baihua Xiao. Degraded document image binarization using structural symmetry of strokes. *Pattern Recognition*, 74:225–240, 2018.

[5] Bhabatosh Chanda Mayank Wadhwani, Debapriya Kundu. Old handwritten document restoration based and clean text extraction using deep learning. *2nd Workshop on Digital Heritage (WDH' 18) IIIT Hyderabad, Hyderabad.*

[6] Reza Farrahi Moghaddam and Mohamed Cheriet. Rsldi: Restoration of single-sided low-quality document images. *Pattern Recognition*, 42(12):3355–3364, 2009.

[7] Ranjan Mondal, Sanchayan Santra, and Bhabatosh Chanda. Dense morphological network: An universal function approximator. *arXiv preprint arXiv:1901.00109*, 2019.

[8] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979.

[9] Ioannis Pratikakis, Konstantinos Zagoris, George Barlas, and Basilis Gatos. Icdar2017 competition on document image binarization (dibco 2017). In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, volume 1, pages 1395–1403. IEEE, 2017.

[10] Ioannis Pratikakis, Konstantinos Zagoris, and Basilis Gatos. Iicfhr 2018 competition on handwritten document image binarization (h-dibco 2018). In *2018 16th International Conference on Frontiers in Handwriting Recognition*. IEEE, 2018.

[11] G. X. Ritter and P. Sussner. An introduction to morphological neural networks. In *ICPR*, volume 4, pages 709–717 vol.4, August 1996.

[12] David Rivest-Hénault, Reza Farrahi Moghaddam, and Mohamed Cheriet. A local linear level set method for the binarization of degraded historical document images. *International Journal on Document Analysis and Recognition (IJDAR)*, 15(2):101–124, 2012.

[13] Jaakko Sauvola and Matti Pietikäinen. Adaptive document image binarization. *Pattern recognition*, 33(2):225–236, 2000.

[14] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 13(4):600–612, 2004.