EasyChair Preprint

№ 8568

Binarization of Color Document Image Based on Adversarial Generative Network and Discrete Wavelet Transform

Yu-Shian Lin, Ting-Yu Lin, Jen-Shiun Chiang and Chih-Chia Chen

August 3, 2022

2022 IET International Conference on Engineering Technologies and Applications (IET ICETA)

Binarization of Color Document Image Based on Adversarial Generative Network and Discrete Wavelet Transform

Yu-Shian Lin
Department of Electrical and Computer Engineering
Tamkang University
New Taipei City, Taiwan
abcpp12383@gmail.com

Ting-Yu Lin
Department of Engineering Science
National Cheng Kung University
Tainan City, Taiwan
tonylin0413@gmail.com

Jen-Shiun Chiang
Department of Electrical and Computer Engineering
Tamkang University
New Taipei City, Taiwan
jsken.chiang@gmail.com

Chih-Chia, Chen
Department of Electrical and Computer Engineering
Tamkang University
New Taipei City, Taiwan
crystal88irene@gmail.com

***Abstract*—Document binarization is an important task to separate the foreground text information in the document image from the background, which is generally applied to the digital archive of historical documents. This paper proposes to use the generative adversarial network for training with a small amount of data. In the first stage, discrete wavelets are used for image enhancement of four-channel images, and local binarization and global binarization are trained separately to obtain the final result in the second stage. The experimental results show that our proposed method has better performance than the classical algorithm on the DIBCO dataset.**

Keywords—Document binarization, generative adversarial network, discrete wavelet transform.

I. Introduction

Document binarization is an issue with a long history of research, and is generally used in the digital collection of historical documents. Due to the age of historical documents, various types of degradation, stains, yellowing, ink oozing and other reasons affect the quality of document binarization. In order to improve the effect of subsequent analyses of the document image, it is an important task to separate the foreground text information in the document image from the background. In recent years, with the rise and development of deep learning technologies, breakthroughs have been made in applying deep learning technologies to computer vision (CV) tasks, including file binarization. Most researches use convolutional neural network (CNN) for semantic segmentation that is based on the pixel-level category prediction of an input image. Usually, for semantic segmentation training, a file image must correspond to a ground truth image, and generally it is not easy to find such a pair of file images. In the public file image database, most of the file images do not correspond to their own ground truth images. Therefore, in the case of insufficient training data, we have to use data augmentation to solve this problem.

II. Related work

In this research we use several techniques, including U-Net[1], GAN[2], EfficientNet[3] and PatchGAN[4]. A brief of introduction of these techniques is discussed as follows.

U-Net

In the ISBI cell tracking competition in 2015, U-Net won the first place in several projects, which mainly solved the segmentation task at the cell level. Our architecture uses U-Net as the encoder for the GAN generator. U-Net forms thicker features by connecting a number of channels, and can train the model with a small number of data.

Generative Adversarial Network (*GAN*)

The Generative Adversarial Network (GAN) is composed of a generator and a discriminator. The input data enters the generator network to produce results corresponding to the output. The purpose is to make the generator network learn fake sample data.

EfficientNet

Proposed by the Google Research Brain Team in 2019, EfficientNet studies model scaling and confirms that balancing network depth, width, and resolution can lead to better performance. The method uniformly scales depth, width, and resolution using a simple but efficient composite factor for all dimensions.

PatchGAN

PatchGAN is a Markov discriminator. Currently Markov discriminators are used in GAN networks such as Pix2Pix and CycleGAN. From PatchGAN, it is completely composed of convolution layers, and the final output is an $n \times n$ matrix. The mean of the output matrix is simply True or False.

III. Proposed Method

We propose a document image binarization method based on generative adversarial network (GAN), which is improved from the architecture of [5]. The overall architecture is divided into two stages, as shown in Fig. 1. In the first stage, the multi-channel GAN neural network is used to perform wavelet transformation to extract the LL frequency band, and then the background information is removed from the local image block, and the color foreground information is extracted. The second stage uses a multi-scale GAN neural network to generate the local binarization result image and the global binarization result image of the document image.

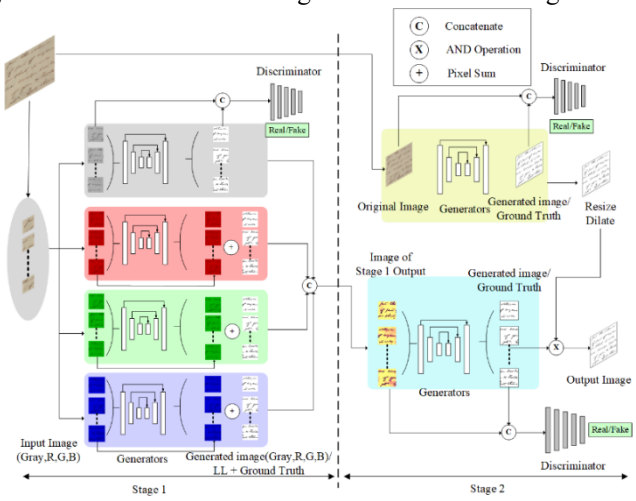

Fig. 1. Main architecture.

The output image of the first stage is a color image, and we would like to remove the background and retain the color information of the foreground. The method is to separate the original image into three channels, then perform discrete wavelet transformation, take out the converted LL frequency band as a characteristic image, and train the image. Since there are four channels to perform training at this stage, the database only provides ground truth of the single-channel images, three additional channels of ground truth need to be generated here. The method is to use the ground truth of the single-channel image to perform bitwise AND operation with the other three-channel LL-band images, respectively, and then generate the binarized image according to the global threshold. After training, the images generated by the four generators are fused to obtain a four-channel color map. The second stage is further divided into local binarization network and global binarization network, and finally the outputs of the two are fused as the final output. Combining the different functions of the two, the output of the two is fused as the final output. Both use the same ground truth. The input of local binarization is divided into several small image blocks, while the input of global binarization does not need to be divided into small image blocks. Both stages use GAN for training, and in the encoder in the generator, we use the U-Net architecture. The encoder of U-Net extracts the image context feature data, and the decoder combines the up-sampling features and the low-dimensional features of the down-sampling stage through skip connections. In the binarization research, U-Net is widely used to improve the performance of the network, and it is in line with the characteristics of less data and not easy to overfit. We adopted EfficientNet as the encoder in the generator and PatchGAN as the discriminator.

IV. Experimental Results

The proposed method is evaluated on degraded document images using the document image binarization datasets DIBCO 2009, H-DIBCO 2010, DIBCO 2011, H-DIBCO 2012, DIBCO 2013, H-DIBCO 2014, and H-DIBCO 2016. A total of 86 downgraded file images were used. We constructed training and test sets from the DIBCO dataset. We selected 34 images from DIBCO 2009, H-DIBCO 2010, H-DIBCO 2012, and Persian Heritage Image Binarization Dataset (PHIBD), and a total of 109 images from the Synchronized Multispectral Ancient Document Image (SMADI) dataset and the Bickley Diary dataset for training. A total of 52 images from DIBCO 2011, DIBCO 2013, H-DIBCO 2014 and H-DIBCO 2016 were used for testing. This paper adopts four evaluation metrics: F-measure (FM), pseudo-F-measure (p-FM), peak signal-to-noise ratio (PSNR), and reciprocal distance distortion (DRD) metrics.

The following are the results of evaluating file image binarization on the DIBCO dataset. The method proposed in this paper is compared with Otsu, Niblack, Sauvola, Vo *et al*., He *et al*., and Zhao *et al*., as shown in Tables I.-IV.

TABLE I. DIBCO 2011 Method and Data Comparison.

Methods	FM	p-FM	PSNR	DRD
Otsu	82.10	85.96	15.72	8.95
Niblack	70.44	73.03	12.39	24.95
Sauvola	82.35	88.63	15.75	7.86
Vo *et al*.	92.58	94.67	19.16	2.38
He *et al*.	91.92	**95.82**	19.49	**2.37**
Zhao *et al*.	**92.62**	95.38	19.58	2.55
Ours	87.71	90.11	**19.64**	3.47

TABLE II. DIBCO 2013 Method and Data Comparison.

Methods	FM	p-FM	PSNR	DRD
Otsu	80.04	83.43	16.63	10.98
Niblack	71.38	73.17	13.54	23.10
Sauvola	82.73	88.37	16.98	7.34
Vo *et al*.	93.43	95.34	20.82	2.26
He *et al*.	93.36	**96.70**	20.88	2.15
Zhao *et al*.	93.86	96.47	21.53	2.32
Ours	**94.88**	96.19	**22.32**	**1.95**

TABLE III. DIBCO 2014 Method and Data Comparison.

Methods	FM	p-FM	PSNR	DRD
Otsu	93.62	95.69	18.72	2.65
Niblack	86.01	88.04	16.54	8.26
Sauvola	83.72	87.49	17.48	5.05
Vo *et al*.	95.97	97.42	21.49	1.09
He *et al*.	95.95	**98.76**	21.60	1.12
Zhao *et al*.	96.09	98.25	21.88	1.20
Ours	**96.88**	98.03	**22.68**	**0.89**

TABLE IV. DIBCO 2016 Method and Data Comparison.

Methods	FM	p-FM	PSNR	DRD
Otsu	86.59	89.92	17.79	5.58
Niblack	72.57	73.51	13.26	24.65
Sauvola	84.27	89.10	17.15	6.09
Vo *et al*.	90.01	93.44	18.74	3.91
He *et al*.	91.19	95.74	19.51	3.02
Zhao *et al*.	89.77	94.85	18.80	3.85
Ours	**91.49**	**96.46**	**19.68**	**2.92**

V. Conclusion

In this paper, four channels are used for wavelet transformation, and the generated images are fused after training different channels respectively. Finally, in order to realize the binarization of the file image, the local binarization network and the global binarization network are combined to balance the extraction of foreground text and background. The experimental results show that the evaluation results on the DIBCO dataset, F-measure (FM), pseudo-F-measure (p-FM), peak signal-to-noise ratio (PSNR) and reciprocal distance distortion (DRD) metrics have excellent results.

References

[1] Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *International Conference on Medical image computing and computer-assisted intervention*, Nov. 2015, pp. 234-241.

[2] I. Goodfellow *et al*, "Generative adversarial nets advances in neural information processing," *International Conference on Neural Information Processing Systems*, vol. 2, pp. 2672-2680, 2014.

[3] M. Tan and Q. Le, "EfficientNet: rethinking model scaling for convolutional neural networks," *International Conference on Machine Learning*, June 2019, pp. 6105-6114.

[4] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *IEEE Conference on Computer Vision and Pattern Recognition*, July 2017, pp. 1125-1134.

[5] S. Suh, J. Kim, P. Lukowicz, and Y. O. Lee "Two-stage generative adversarial networks for document image binarization with color noise and background removal," *Pattern Recognition*, vol. 130, pp. 13, 2022.