ISSN 2616-5775 Vol. 7, Issue 2: 19-25, DOI: 10.25236/AJCIS.2024.070203

A scene-text feature enhanced halftone method based on invertible neural network

Zhang Huajian[1,a], Mu Dazhong[1,b,*]

[1]Beijing Key Laboratory of Signal and Information Processing for High-end Printing Equipment, Beijing Institution of Graphic Communication, Beijing, China
[a]woshizhanghuajian@hotmail.com, [b]dazhongmu@bigc.edu.com
[*]Corresponding author

Abstract: *In this paper, we present a halftone method based on invertible neural network (INN). We use the spectral features that halftone image dithering should have as constraints during training to allow the network to learn how to generate satisfactory halftone images. We use the original grayscale image as input during training, and use the halftone image generated through traditional error diffusion methods as the reference target. Our method outputs halftone images that preserve textual information that was not preserved in the error diffusion halftone method, enhancing the direct readability of the textual content that appears in the halftone images.*

Keywords: *Halftone, Dithering, Invertible Neural Network, Deep Learning*

1. Introduction

A halftone image is a binary image consisting of black and white dots. Halftone technique is a technique to convert the original continuous tone image into halftone image. It is different from the common thresholding method that simply divides the image into black and white parts according to the threshold value, which will make the tone of the image almost disappear. The halftone method dithers the black and white points to reflect the tonal variations in the original image through density changes.

Halftoning is a kind of image transforming technique with a long history that first appeared in the 1980s [1]. A variety of halftone methods were proposed during this time [2] [3]. Halftone techniques can be categorized as periodic or non-periodic depending on the type of pattern, and as clustered or discrete depending on the type of "dot". The widely used error-diffusion (ED) halftone method [4] is a non-periodic, dispersed technique. ED halftones are popular because of their relative simplicity and good quality of blue noise generation.

Blue noise is considered to be the central element that makes halftone images visually appealing [5]. Noise is described by the name of a color according to its frequency. White noise is the noise that has a uniform distribution of power at different frequencies and blue noise is the noise that has power concentrated in the high-frequency region. The noise in the halftone map generated by the ED halftone technique is mainly concentrated in the high-frequency region, while the noise in the low-frequency region accounts for a small percentage. According to the theory of Human Vision System, the human eye can be equated to a low-pass filter, which can filter out the high-frequency blue noise, so the halftone image in the human eye can be approximated to the original continuous tone image.

In the task of halftoning natural images, error diffusion halftoning achieved good results. However, as shown in Figure 1, when text appears in the natural scene image, the ED method fails to generate text that can be read directly, but drowns the textual information in noise, so we propose a halftone method that can enhance textual features.

The paper is organized as follows, in section 1 we introduced the background of halftone techniques and the problems encountered by existing halftone methods. In section 2 we will describe the specific implementation of our algorithm. Part 3 will be the evaluation of the image quality results generated by the proposed method. Finally in part 4, we summarize the proposed method.

ISSN 2616-5775 Vol. 7, Issue 2: 19-25, DOI: 10.25236/AJCIS.2024.070203

(a) *(b)*

Figure 1: The ED method failed to obtain the text from the original image. (a) grayscale image. (b) halftoning image of (a).

2. Proposed Method

The Invertible Neural Network (INN) is a type of neural network that was proposed to solve ill-posed problems. Laurent et al. first introduced the prototype of INN in 2014 [6][7]. Traditional neural networks like CNN often lose critical information in the forward process, making it infeasible to recover the input data of a system from measurable results. INNs can easily overcome it with its specially designed features. Firstly, the mapping from input to output is a bijective function, meaning that its inverse exists. Both forward and inverse mappings can be efficiently computed, and both mappings have a processable Jacobi matrix that allows explicit computation of the posterior probabilities. These features ensure that in ideal environment information can be fully preserved in both forward and inverse calculations. The primary goal of INNs is to solve the ill-posed inverse problem, where a single observation may have multiple corresponding inputs. In the image halftone process, there is a quantization process, which is a many-to-one arithmetic process that quantizes N-level tones to 0 or 1. Therefore, using INNs to solve the problem of halftone generating is reasonable. In recent times, INNs have been utilized to great effect in various image processing tasks such as invertible image scaling [8], invertible image decolorization [9], invertible inverse halftoning [10], and invertible image signal processing [11], with remarkable outcomes.

2.1. Network Structure

Our proposed method is an invertible technique for halftoning on grayscale images. We also expect that the grayscale texture information lost in the halftoning process can be retained in the inverse halftone image, which means that the original image can be recovered from the halftone image. Therefore, we try to use INN to encode the grayscale information into the halftone image. In a general INN model, people usually choose to hide what is to be encoded in the output latent variables conforming to a Gaussian distribution. However, Xia et al. [12] identified a problem that would be encountered when generating halftone maps directly with neural networks, flatness degradation, which would result in the inability of jitter patterns to appear in regions of the output image that have a constant gray scale. They proposed to use Noise Inception Block (NIB) to solve this problem by adding a channel of Gaussian noise to the input to artificially introduce space domain variations to the constant regions. We therefore made changes to the INN model structure as in Figure 2. We changed the inputs from two identical pictures to an input picture and Gaussian noise, and the output picture and Gaussian latent variables to output pictures for both channels.

So we can easily get the formulation of the full model:

$$(I^1_{p-HT}, I^2_{p-HT}) = f^1(I_{GS}, I_{GN}) \tag{1}$$

$$I^1_{HT} = Bin(I^1_{p-HT}) \tag{2}$$

$$(I^{rec}_{GS}, I^{rec}_{GN}) = f^{-1}(I^1_{HT}, I^1_{HT}) \tag{3}$$

Where I^1_{p-HT} and I^2_{p-HT} denote the pseudo halftone images of the two channels without binarization, I_{GS} and I_{GN} are respectively the grayscale images and the added Gaussian noise. f^1 and f^{-1} are the forward and inverse transformations of the network. I^1_{HT} denotes the binarized halftone image and $Bin(\cdot)$ denotes the binary gate. I^{rec}_{GS} and I^{rec}_{GN} denote the grayscale images and Gaussian noise recovered by the inverse transformation of the network. The input to the inverse transform in Eq.

ISSN 2616-5775 Vol. 7, Issue 2: 19-25, DOI: 10.25236/AJCIS.2024.070203

(3) uses only I_{HT}^{1} because it is impractical to set up two halftone images at the same time, so only one of them can be used for recovery.

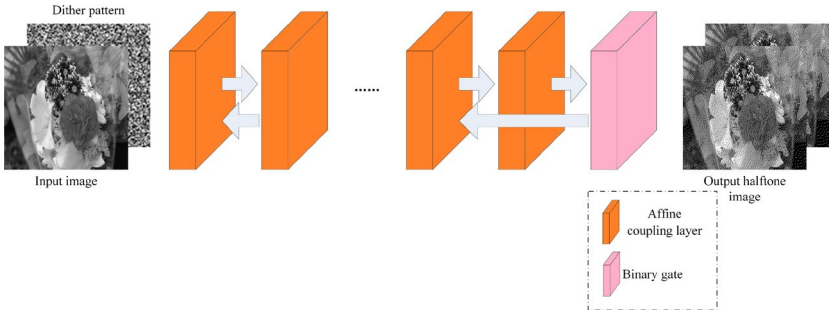

Figure 2: The network structure of the proposed network

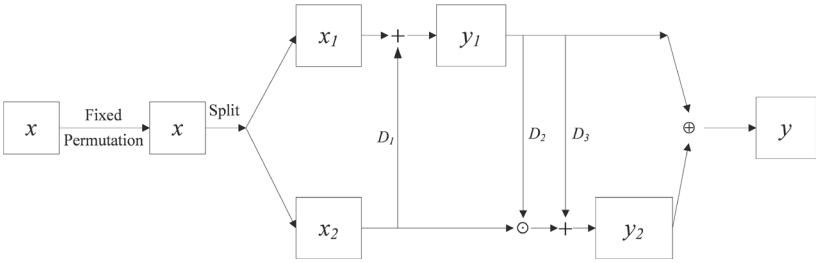

Figure 3: Structure of the Affine Coupling Layer

INN is composed of affine coupling layers (ACL). The ACL is the core of achieving network invertibility. The structure of the ACL makes it possible to obtain the original inputs accurately in the reverse computation. The bilateral coupling layer can process the input x_1 and x_2 at the same time. Figure 3 shows the detailed structure of the bilateral coupling layer. The formulas for the output y are shown as follows

$$y_1 = x_1 + D_1(x_2) \tag{4}$$

$$y_2 = x_2 \odot D_2(y_1) + D_3(y_1) \tag{5}$$

$$y = y_1 \oplus y_2 \tag{6}$$

Similarly, their inverse can be expressed as

$$x_2 = (y_2 - D_3(y_1)) \odot (1 / D_2(y_1)) \tag{7}$$

$$x_1 = y_1 + D_1(x_2) \tag{8}$$

$$x = x_1 \oplus x_2 \tag{9}$$

Where D_1, D_2, D_3 are three mapping transformation functions, they can be any functions, we choose to use five-layer Convolutional Neural Networks. $\odot$ denotes to Hadamard product and $\oplus$ denotes to concatenation along the channel axis.

2.2. Loss Function

Following the steps of Xia et al. [12], our loss function is made of three parts: forward loss, recovery loss and special loss.

$$L = \alpha_1 L_{forward} + \alpha_2 L_{rec} + \alpha_3 L_{special} \tag{10}$$

The forward loss is the loss function in the forward direction

$$L_{forward} = \beta_1 L_{quan} + \beta_2 L_{tone} + \beta_3 L_{channel} + \beta_4 L_{guidance} \tag{11}$$

ISSN 2616-5775 Vol. 7, Issue 2: 19-25, DOI: 10.25236/AJCIS.2024.070203

$$L_{quan} = \left\| I_{HT}^{1} - I_{p-HT}^{1} \right\|_1 \tag{12}$$

$$L_{tone} = \left\| F(I_{HT}^{1}) - I_{GS} \right\|_1 \tag{13}$$

$$L_{channel} = \left\| I_{p-HT}^{1} - I_{p-HT}^{2} \right\|_1 \tag{14}$$

$$L_{guidance} = \left\| IHT(I_{HT}^{1}) - I_{GS} \right\|_1 \tag{15}$$

Where F denotes low-pass filtering and IHT is the inverse halftone processing using a pre-trained model. Quantization loss is used away to measure the gap between the direct output of the network and the binarized result. Tone loss is used to measure the difference in grayscale between the halftone image and the original image. Channel loss is used to make the two channels of the network output as similar as possible. Guidance loss is used to process the output halftone image with a pre-trained IHT network, which allows the network to better capture the halftone pattern.

We use the recovery loss to describe the gap between the gray-scale image recovered by the network and the original image, which can be written as

$$L_{rec} = \left\| I_{GS}^{rec} - I_{GS} \right\|_1 \tag{16}$$

In order to make the network better learn generating dither patterns that satisfy the quality of blue noise, we train the network so that for every realistic image we process, we also generate a halftone image of a uniform picture with random grayness level. The loss function for this part can be written as

$$L_{special} = \left\| DCT(I_{HT}^{1}) - I_{GS} \right\|_1 \tag{17}$$

Where DCT denotes a discrete cosine transform of the picture, which aims to discard the high-frequency regions of the picture and keep only the low-frequency regions. Constraining it will place the noise as much as possible in the high-frequency region and satisfy the blue noise requirement.

In the above equations, α_1 to β_4 are the coefficients. Specifically, in our training they are set as α_1=1, α_2=0.1, α_3=1, α_1=0.3, β_1=0.25, β_2=0.6, β_3=0.1, β_4=0.3.

3. Experiment

3.1. Implement Detail

We conducted our experiments on the DIV2K dataset [13], which was originally designed for the image super-resolution task. The dataset comprises 900 high-resolution images with 2k resolution, which are categorized into realistic landscapes, portraits, and street scenes. We used 800 images for training, and 100 for testing.

To generate the reference halftone images, we used the Floyd-Steinberg method and paired them with the original grayscale images. To accelerate training, every time we put an image into the network, we randomly cropped it into small block of 256×256 resolution. The full-resolution images were used in the testing stage.

We executed our network model using PyTorch on a Nvidia Tesla P40 with 16 GB memory. We chose the Xavier initializer and Adam optimizer. We started the training with a learning rate of 1×10^{-4} and halved it at 15, 40, and 100 epochs. The batch size was set to 1, and the number of affine coupling layers was set to 11. The entire training process lasted for 200 epochs.

3.2. Experiment Result

In this section, we test our model on the power spectral density maps and also compare the differences between our model and the halftone images generated by existing methods. Firstly, the halftone maps of our uniform grayscale images and their power spectral density maps are shown in Figure 4. According to Eq. (18), the principal frequencies of these three shades of grayness are 0.30, 0.59 and 0.66, and from Figure 4 we can see that our model better places the noise power peaks near the principal frequency,

ISSN 2616-5775 Vol. 7, Issue 2: 19-25, DOI: 10.25236/AJCIS.2024.070203

leaving only little noise in the low-frequency part.

$$f_{principal} = \begin{cases} \sqrt{g/255} & , g \le 128 \\ \sqrt{(255-g)/255} & , g > 128 \end{cases} \tag{18}$$

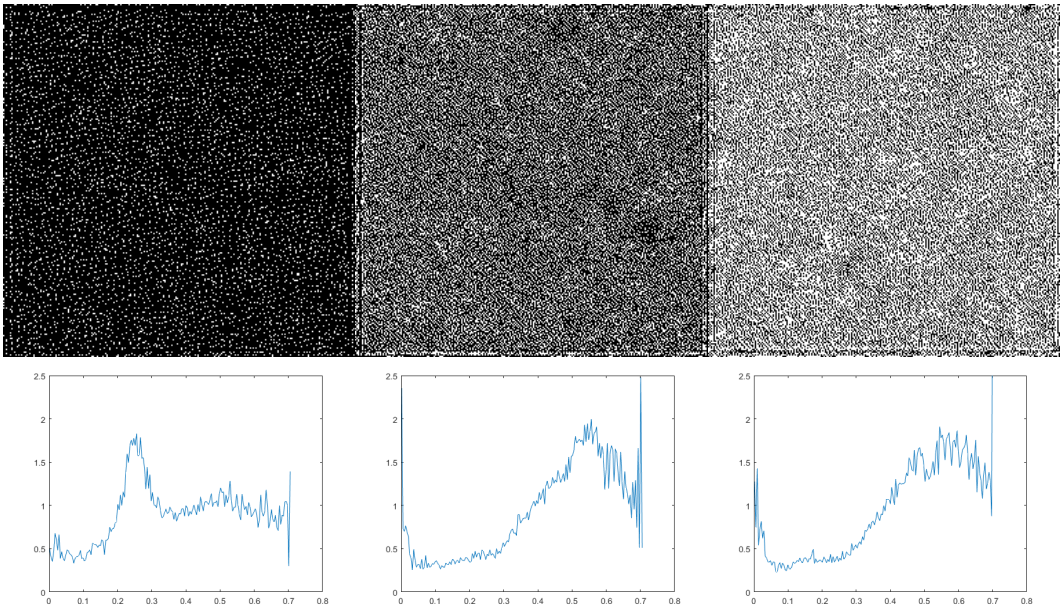

Figure 4: The top row shows the generated halftone images of a series of constant grayness. The bottom row shows the power spectral density map of the corresponding images. From left to right, their grayscale values are 23, 89 and 144 out of 255. The horizontal axis in the following row of pictures represents the radial frequency and the vertical axis represents the power spectral density.

We use the images in the DIV2K test set to verify our model. We use the Floyd-Steinberg error diffusion (FSED) method and the Reversible Binary Pattern (RBP) [12] method to compare with the proposed method, and the results are shown in Figure 5. From Figure 5, it can be seen that the proposed method is indeed insufficient compared to these well-established methods in the face of landscape and portrait images, but when text content that requires a high degree of readability appears in the image, our method can better retain this information, while both FSED method and RBP method failed to do so.

4. Conclusion

In this paper, we proposed an INN based halftone method. By encouraging the network to generate blue noise during the training process, we can generate appropriate halftone images. According to the experimental results, we found that the proposed method achieved good results on natural scene images with text content, enhancing readability of text.

ISSN 2616-5775 Vol. 7, Issue 2: 19-25, DOI: 10.25236/AJCIS.2024.070203

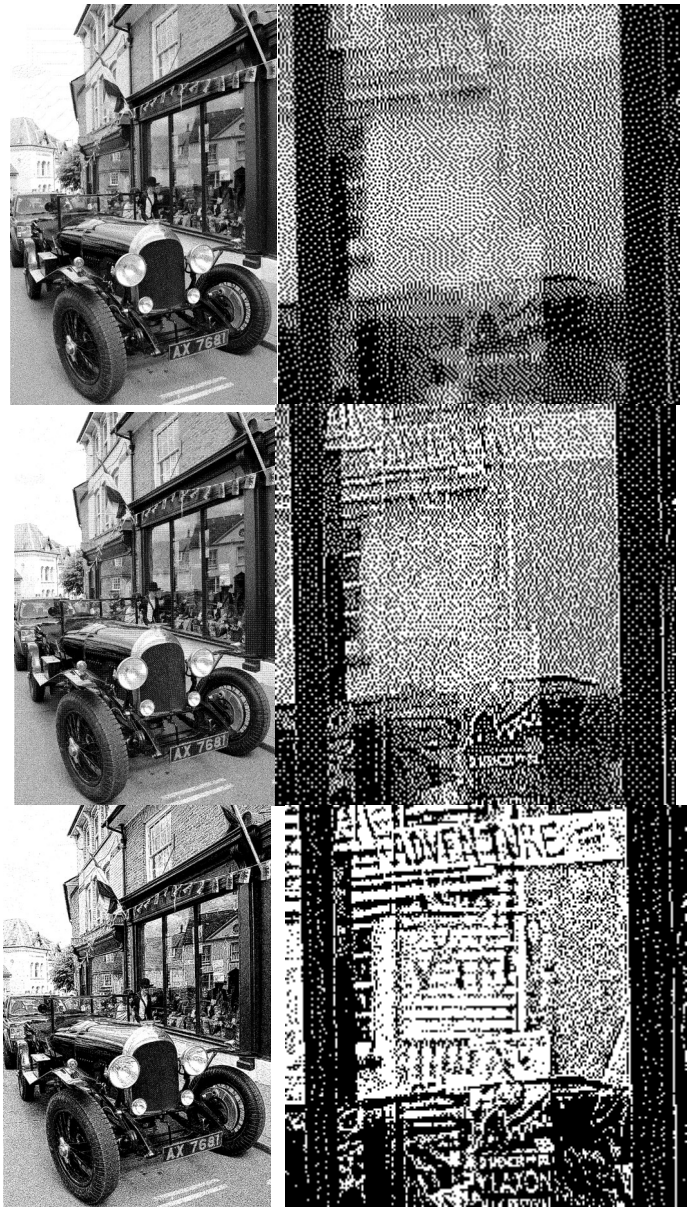

Figure 5: The comparison between the proposed method, FSED and RBP. From top to bottom, they are the grayscale image, FSED halftone image, RBP halftone image and our image. The second column zooms the area with text.

References

[1] Ulichney, Robert. Digital halftoning[M]. MIT Press, 1987.
[2] Knuth DE. Digital halftones by dot diffusion[J]. ACM Transactions on Graphics (TOG). 1987, 6(4): 245-73.
[3] Chang J, Alain B, Ostromoukhov V. Structure-aware error diffusion[C]. InACM SIGGRAPH Asia 2009 papers (pp. 1-8), 2009.
[4] Floyd, R. W. An adaptive algorithm for spatial gray-scale[C]. In Proc. Soc. Inf. Disp. (Vol. 17, pp. 75-77), 1976.
[5] Pang W M, Qu Y, Wong T, et al. Structure-aware halftoning[C]. In ACM SIGGRAPH 2008 papers (pp. 1-8), 2008.
[6] Laurent Dinh, David Krueger, Yoshua Bengio, NICE: Non-linear Independent Components Estimation[C]. In ICLR workshop 2015
[7] Laurent Dinh, Jascha Sohl-Dickstein, Samy Bengio, Density estimation using Real NVP[C]. in ICLR 2017.
[8] Xiao Mingqing, et al. Invertible image rescaling[C]. In Computer Vision–ECCV 2020: 16th

ISSN 2616-5775 Vol. 7, Issue 2: 19-25, DOI: 10.25236/AJCIS.2024.070203

European Conference 2020, Proceedings, Part I 16.
[9] R Zhao, T Liu, J Xiao, et al. Invertible Image Decolorization[J]. IEEE Transactions on Image Processing, 2021, vol. 30, pp. 6081-6095.
[10] Zhang Huajian, Mu Dazhong and Cao Peng, An Inverse Halftoning Method Using Invertible Neural Network[C]. 2023 8th International Conference on Intelligent Computing and Signal Processing, 2023.
[11] Y. Xing, Z. Qian and Q. Chen, "Invertible Image Signal Processing[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.
[12] Xia Menghan, et al. Deep halftoning with reversible binary pattern[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.
[13] E Agustsson and R Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017.