Reconstruction of Multitone BTC Images using Conditional Generative Adversarial Nets

Jing Ming Guo* and Sankarasrinivasan S†
*†National Taiwan University of Science and Technology
* E-mail: jmguo@seed.net.tw Tel: +886-27303241
†E-mail: D10507805@mail.ntust.edu.tw Tel: +886-978374981

***Abstract*—** Multitone Block Truncation Coding (MT-BTC) image is the superior version of halftone based BTC compression methods. The MT-BTC images are developed based on ordered dithering method which utilize multitone dithering mask for image construction. During the transformation, the original image is processed in a block-wise manner and is replaced in terms of the maximum, minimum and their intermediate values of the respective block. In comparison with standard compressions techniques, MT-BTC possess very unique representation and suffers from inherent halftone noises. In this paper, a simplified version of image to image translation architecture is developed based on cGAN's. To begin with, a MT-BTC database is developed using the latest multitone approach, and it comprise of around 10,000 images. Further, the proposed cGAN's model is optimized to perform with minimal layers and reduced parameters. The PatchGAN discriminator is adjusted to judge over the patch size of 64x64 which has good impact over quality improvements. From the comprehensive performance evaluation, it has been validated that the proposed approach can achieve consistent and improved reconstruction quality.

I. Introduction

Digital halftoning [1] [2] is a process of transforming a gray image into its approximate binary version which are useful in printing applications. On the other hand, block truncation coding is a technique to compress a gray scale image in a block wise manner and the processed image block is represented in terms of high and low mean along with its bitmap image. Hence the compressed form for a block consist of a bitmap image and two integer values as shown in Fig. 1.

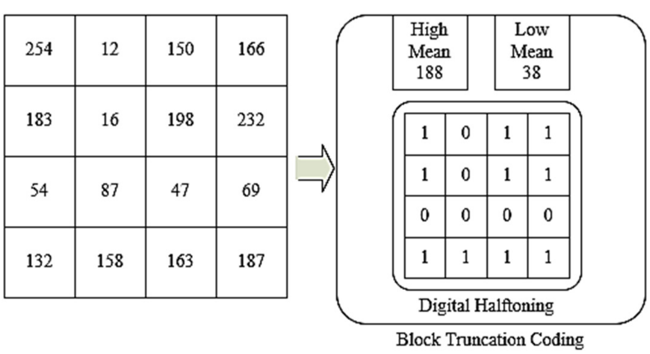


Fig. 1. Digital halftoning vs Block truncation coding

From close observation it can be inferred that both halftoning and BTC method attempts to obtain the binary version for the whole image and an image block respectively. With this association, many of the halftoning techniques are extended to BTC to reduce its computational complexity and to improve the image quality [4].

In halftone evolution, the Digital multitoning [3] is recently introduced in which a gray scale image is represented with more than two number of tones. For instance, 3-tone image consist of {0, 128, 255}. And as in case of halftones, the digital multitoning techniques are also extended to BTC to obtain MT-BTC technique. The MT-BTC images [5] are observed to have a better image quality with limited compression ratio. The present work deals with reconstruction of MT-BTC images using the deep learning models.

Among the various deep learning models for reconstruction, the image-to-image (I2I) translation algorithm is considered and optimized with respect to the MT-BTC reconstruction. As the I2I model relies on the generative adversarial network (GAN) to learn the loss function, the method is observed to have reconstruct image with better image quality. The I2I model is based on the conditional GAN which requires that the network need to be trained with the with a MT-BTC image along with the noise.

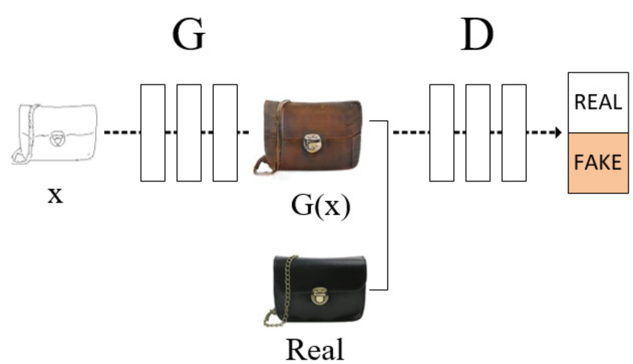


Fig. 2. Conditional Generative Adversarial Networks (cGAN's)

From the literature, it has been validated that the cGAN's are very effective in reconstruct the image from the edge information and various style transfer. But the image reconstruction case is quite easy to perform as the approximate information is available. Hence in the present study the architecture is optimized in this aspect.

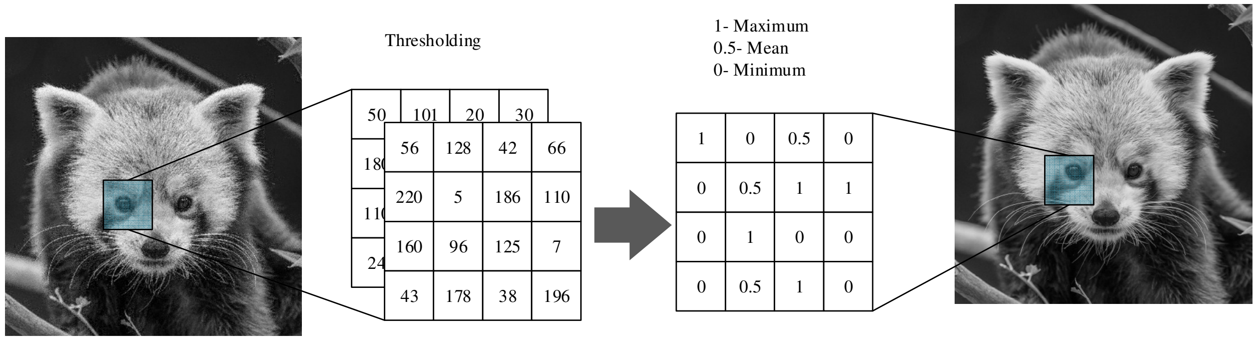


Fig. 3.MT-BTC Image Construction for 3-tones

Existing methods for the reconstruction of halftone based BTC images relies on hand crafted filters or codebooks to improvise the image quality. The main problem associated with the halftone BTC images are the inherent halftone impulse noise. Guo. et. al [6] proposed an optimized classified filter to effective remove the impulse noises for the ordered dithering and error diffusion BTC images. Another strategy involves vector quantization and sparsity based solution for effective image reconstruction.

The image quality associated with the different block size of the MT-BTC is shown in Table I. It has been inferred that the image quality degrades consistently with respect to block size, but at the same instance, it results in better compression ratio. Hence, MT-BTC image with higher block size are preferable to obtain a better compression ratio.

TABLE I. MT-BTC IMAGE QUALITY FOR DIFFERENT BLOCK SIZE.

Block Size	MT-BTC
8	0.9704
16	0.9664
32	0.9340
64	0.8891

Fig. 4. MT-BTC Image with blocking effects and impulse noise.

From Figure 4, it can be observed that the MT-BTC images with higher block sizes suffers from blocking effects and impulsive noises. As the hand crafted based restoration techniques are not very effective to eliminate such artifacts, a deep learning based strategy is proposed to handle the issue.

The important contribution of the paper is as follows:

- a MT-BTC database is constructed for the multitone BTC images constructed with different block sizes ranging from 0 to 64. The database comprises of 10,000 MT-BTC images.
- A simplified conditional GAN is proposed with reduced convolutional filters and layers. This results in fast training and significant reduction of parameters.
- The loss function of the existing architecture is also altered relevant to the image reconstruction problem. The standard cGAN's typically involves L1 loss, but for the MT-BTC reconstruction it has been removed and GAN is trained based on the loss computed over the effective image patches.

II. MULTITONE BLOCK TRUNCATION CODING IMAGE

Figure 3 illustrates the methodology involved with the MT-BTC construction. It can be observed that the original gray scale image has to undergo thresholding using many dithering screens and the output values is assigned with multitone data. In this example, the values {1, 0.5, 0} corresponds to different tonal information and it can be interpreted as white, middle-gray and black in terms of printing and for compression they are assigned with maximum, average of maximum and minimum term and the minimum value of the block. To obtain the best image quality the dispersed dot dithering screens is used in default.

$$MT = \begin{cases} 1 & if\ \ org\,[p,q] > DA^1; \\ 0.5 & if\ DA^1 \geq org\,[p,q] > DA^2; \\ 0 & otherwise. \end{cases} \tag{1}$$

In Eq. 1, the MT refers to the multitone BTC image, org refers to the original image and the DA^1 and DA^2 refers to the dithering screen.

III. MT-BTC Image Reconstruction

Image to image translation [8] often deals with the image classification, reconstruction or regression problem that operates in a per-pixel level. As the input and output pixels are mapped in one to one basis, they are assumed to be conditionally independent over each other. Usually the network relies on two loss function such as L_1 and a structural loss element. The main role of the conditional GANs is to learn the structural loss L_{cGAN} during the training which enhances the realistic nature of the reconstructed image.

The conventional generator networks deal with mapping a random noise vector to an expected output image. On the other hand, Condition GAN learn the mapping between an observed image x and a random noise vector z to the output image. Relevantly, the objective function can be defined as

$$L_{cGAN}\,(G,D) = E_{x,y}[\log D(x,y)] + E_{x,z}[\log(1 - D\big(x,G(x,z)\big)] \quad (2)$$

During the optimization, the G tries to minimizes the function whereas the D tries to maximize the function. In addition to this, integration of L_1 or L_2 loss is also proven to be an effective strategy and the final objective function is provided as

$$G^* = \arg min_G\, max_G\, L_{cGAN}\,(G,D)+\lambda\mathcal{L}_{L1}(G) \quad (3)$$

The added output noise vector is of gaussian type.

Usually the L_1 or L_2 loss mainly suppress the high frequency elements, and hence the image output usually is blurred in nature. In general, the model rely on the L_1 loss to accurately capture the low frequency information and restricting the GAN architecture to only to the high frequency information that is related to the structural information in the image patch.

The discriminator operates as a Patch-GAN which penalizes the structural information at the scale of the patches, and the discriminator tries to determine whether the output image patch of size NxN is real or fake. In this approach, the image pixels are treated as a Markov random field elements and the assumption regarding the independency between the pixels is established based on the patch diameter.

The generator model comprises of an encoder-decoder architecture which consist of Convolutional layers are used along with the batch normalization technique. The ReLu layers is used as the activation function layer. The filter size of the convolution layers is 4x4 with the stride rate of two.

The number of filters for each convolutional layer of the encoder and decoder is provided as

Encoder: CON64-CON128-CON256-CON512-CON512

Decoder: CON512-CON512-CON256-CON128-CON64

The number of filters for the discriminator architecture is

Discriminator: CON64-CON128-CON256-CON512-CON512

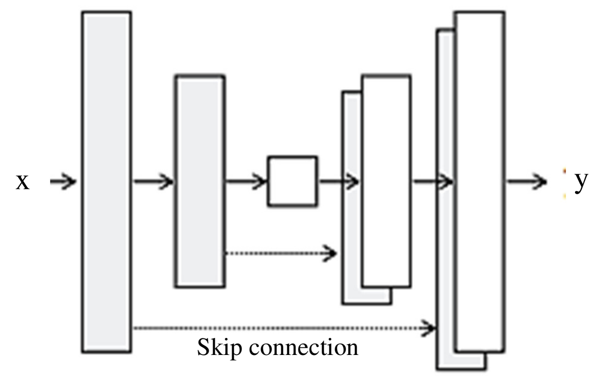


Fig. 5. Encoder-Decoder with skip connection

Figure 5, illustrates the architecture of the generator in which the encoder-decoder is provided with a skip connection. The initial weights are initialized based on the Gaussian distribution with mean 0 and standard deviation 0.01.

As the MT-BTC images are well approximated with respect to the given image, the main challenge in restoration is to remove the inherent impulse noise associated with halftoning and to improvise the structural similarity. From our experiments, it has been inferred that the L1 loss is not contributing significant impact to the visual quality as low pass information is already available and hence the L1 loss term is assigned with less weightage. And as image patches with high structural information are going to play a key role in loss function optimization, the GAN network is optimized considering only the effective patches. The effective patches are estimated using the variance and this also results in faster learning. The discriminator works as a PatchGAN in which it tries to predict the generated image in a patch wise manner. The patch size of the discriminator is set in accordance with the block size over which the MT-BTC image are constructed. As the MT-BTC is block-wise compression technique, the use of PatchGAN is very relevant and found to be very effective in obtaining a good reconstruction.

As the MT-BTC image is obtained by processing the gray scale image in the block wise manner, this Patch-GAN is very relevant and applicable for optimization. For the generation of MT-BTC image various block sizes has been applied such as 8, 16, 32 and 64. During the training, the algorithm is trained for MT-BTC images of different sizes and accordingly the patch size diameter is also changed.

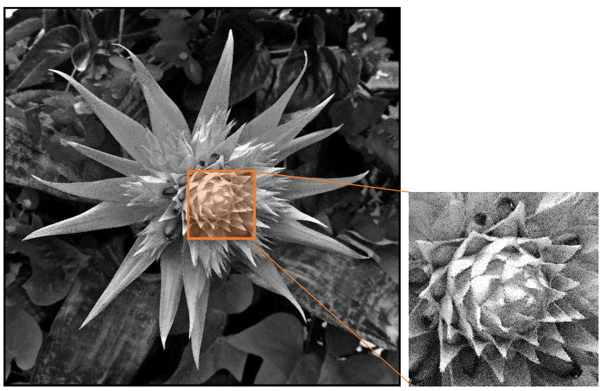
The number of images used for training and validation is 4000 and 1000 images. The algorithm is tested for the 100 epoch and the generator loss is observed to be minimal and stabilized. As the MT-BTC with higher block sizes has more noises, the PatchGAN is set to 64x64 and it helps to reduce the noise significantly.

For quality assessment, structural similarity index (SSIM) is used to evaluate between the original, MT-BTC and its reconstructed version.

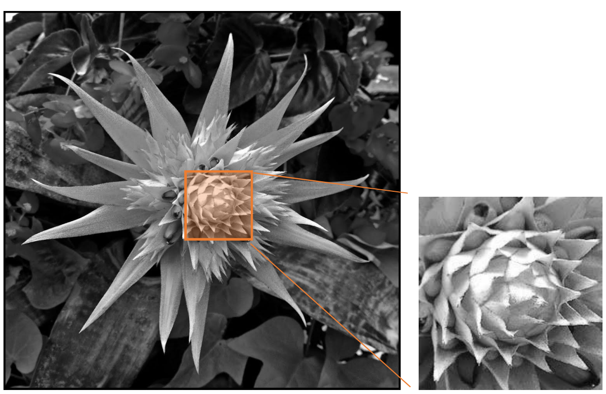
TABLE II. IMAGE QUALITY OF RECONSTRUCTED MT-BTC

Block Size	MT-BTC	Guo. et. al [6]	Reconstructed MT-BTC
8	0.9704	0.9799	0.9816
16	0.9664	0.9712	0.9801
32	0.9340	0.9401	0.9566
64	0.8891	0.9044	0.9233

Table II shows the results of improvements in the image quality of the reconstructed MT-BTC image. It can be noted that the reconstruction rate is superior for the images with higher block sizes and it is due to the presence of heavy visible noises. The proposed method also performs better with respect to other methods.



(a) MT-BTC Image



(b) Reconstructed MT-BTC Image

Fig. 6. Reconstructed MT-BTC Image.

Fig. 6. shows the actual and reconstructed version of the MT-BTC image. It can be clearly seen that the significant amount of impulsive noises present in the existing MT-BTC image is removed to a significant level. And the edge information is very much preserved and accurately presented in the reconstructed image.

IV. CONCLUSION

A deep learning based approach is proposed for the superior reconstruction of the MT-BTC images. To begin with, the MT-BTC database is constructed which comprise of around 10,000 images for training. The inherent noises associated with a MT-BTC images such as blocking artifacts and impulsive noises is studied and Conditional General Adversarial Networks (cGAN's) are exploited to solve this problem. As the image to image translation architecture is formulated for the edge to image and many complex problems, the present work considerably reduced the number of filter and layers with respect to existing work. The loss function of the existing architecture is also fine-tuned to perform for MT-BTC images, and the loss functions that corresponds to retaining low pass information is neglected. The PatchGAN based discriminator model is set to perform for block size of 64x64, and the decision over real or fake image is decided for this specific image size. From the results, it is validated that the proposed strategy obtained very good image quality and the visible inherent halftone based noises are eliminated significantly. In future work, the architecture is extended to perform for reconstruction of other compressed image versions. For dataset and code: https://github.com/SankarSrin

REFERENCES

[1] D. L. Lau, and R. A. Gonzalo., Modern digital halftoning. CRC Press, 2008.

[2] J. M. Guo and S. Sankarasrinivasan. "Digital Halftone Database (DHD): A Comprehensive Analysis on Halftone Type," *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 1091-1099, 2018.

[3] Y. F. Liu et. al., "Clustered-dot screen design for digital multitoning," *IEEE Transactions on Image Processing*, vol. 25, no. 7, pp. 2971-2982, 2016.

[4] J. M. Guo and M. F. Wu., "Improved block truncation coding based on the void-and-cluster dithering approach," *IEEE Transactions on image processing*, vol. 18, no. 1, pp. 211-213, 2008.

[5] J. M. Guo, and S. Sankarasrinivasan, "Enhanced block truncation coding image using digital multitone screen," *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 672-676, Dec. 2017.

[6] J. M. Guo. et. al., "Blocking effect and halftoning impulse suppression in an ED/OD BTC image with optimized texture-dependent filter sets," *International Conference on System Science and Engineering*, pp. 593-596, 2011.

[7] J. M. Guo et.al.," Halftoning-based block truncation coding image restoration," *Journal of Visual Communication and Image Representation*, vol. 35, pp. 193-197, 2016.

[8] Isola, Phillip et. al., "Image-to-image translation with conditional adversarial networks," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1125-1134. 2017.

[9] O. Ronneberger et. al., "U-net: Convolutional networks for biomedical image segmentation," *In International Conference on Medical image computing and computer-assisted intervention*, pp. 234-241, 2015.