2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Low Dose CT Image Denoising Using Boosting Attention Fusion GAN with Perceptual Loss

Luella Marcos[1], Javad Alirezaie[1] *Senior IEEE Member*, and Paul Babyn[2]

***Abstract*—Image denoising of Low-dose computed tomography (LDCT) images has continues to receive attention in the research community due to ongoing concerns about high-dose radiation exposure of patients for diagnosis. The use of low radiation CT image, however, could lead to inaccurate diagnosis due to the presence of noise. Deep learning techniques are being integrated into denoising methods to address this problem. In this paper, a General Adversarial Network (GAN) composed of boosting fusion of spatial and channel attention modules is proposed. These modules are embedded in the denoiser to address the limitations of other GAN-based denoising models that tend to only focus on the local processing and neglect the dependencies of creating feature maps with spatial- and channel- wise image characteristics. This study aims to preserve structural details of LDCT images by applying boosting attention modules, prevents edge over-smoothing by integrating perceptual loss via VGG16 pre-trained network, and finally, improves the computational efficiency by taking advantage of deep learning techniques and GPU parallel computation.**

***Index Terms*—Medical imaging; Computed tomography; Image denoising; Generative adversarial network; Attention fusion**

I. Introduction

The use of X-ray radiation in computed tomography (CT) scans is widely used as an effective tool for medical diagnosis. However, there is an increasing concern about the health risks of the patients when exposed to high radiation like cancer cases due to the induced CT-related X-ray radiation [1]. The research community aims to minimize the exposure of the patients by limiting the dosage of radiation. The quality of the image may be affected due to the presence of noise, and hence, giving inaccurate diagnosis. A solution would be enhancing the image quality by applying efficient image denoising techniques on LDCT images.

Convolutional neural networks (CNN)-based denoising models have been proven to address the issues of spatial domain filtering and variational denoising techniques [6]. The two main types of CNN-based denoising methods include multi-layer perceptron (ML) and deep learning methods such as in [1], [2], and [4]. MLP-based models are often composed of encoders and decoders, which usually follow a feed-forward artificial neural network. Although this method has a better interpretability than the optimization algorithms, the main drawback would be the uncontrollable number of parameters during the denoising process because of the fully connected architecture. Deep learning, however, has proven to be more effective for image processing such as AlexNet [7], VGGNet [8] and ResNet [9]. One problem with deep learning networks is the gradient problem and also, the insufficient evaluation of denoised images during the training process. Generative adversarial network (GAN) - based frameworks utilizes the advantage of GPU's parallel architecture which could solve optimization problems mentioned. Further, a GAN-based denoising model is usually composed of a generator, which generates the desired data, and a discriminator, which is responsible for determining whether the data is from the training set or the generated set of data produced from the generator. Different metrics have been used to avoid the problem of vanishing gradient which commonly occurs in the generator.

Even though the performance of GANs has shown significant development in image processing, there are still room for improvement. In terms of denoising low-dose CT images, GANs struggle to find efficient ways in retaining image information especially the fine details in CT images like blood vessels and small lesions. Park et al. [10] proposed a fidelity-embedded GAN (f-GAN), which uses maximum a posteriori (MAP), a statistical image reconstruction technique for data fitting during noise filtering. Despite of producing accurate results, a drawback would be the time consumed during the training process. In this study, GPU parallel computation would be applied to reduce such computational cost. Further, Yin et al. [11] proposed an integration of multi-perceptual loss (MPL) and fidelity loss into GAN in order to accomplish unpaired image denoising. This framework also focused on maintaining the high-level semantic features of the image by applying MPL. This study has shown how perceptual loss can keep the perceptual features of CT images using the unsupervised learning method.

The main contributions of the proposed model in this paper are summarized as follows:

1) Preservation of structural details of LDCT images by adopting boosting attention modules integrated in GAN based on the framework in [3] .
2) Prevention of edge over-smoothing by integrating the use of perceptual loss using VGG16 network, which was proven to be effective in [5].
3) Improvement of computational efficiency by implementing a deep learning approach instead of an iterative-based methods.

The remainder of this paper is arranged as follows: Section II discusses the proposed denoising model architecture; Section IV presents the quantitative and visual results of the experiment; finally, concluding remarks will be in Section V.

[1]Department of Electrical and Computer Engineering, Ryerson Univeristy, Toronto, ON M5B2K3 Canada (e-mail: lgmarcos@ryerson.ca; javad@ee.ryerson.ca); [2] Department of Medical Imaging, University of Saskatoon Health Region, Royal University Hospital, Saskatoon, SK S7N0W8 Canada (e-mail: paul.babyn@saskatoonhealthregion.ca)

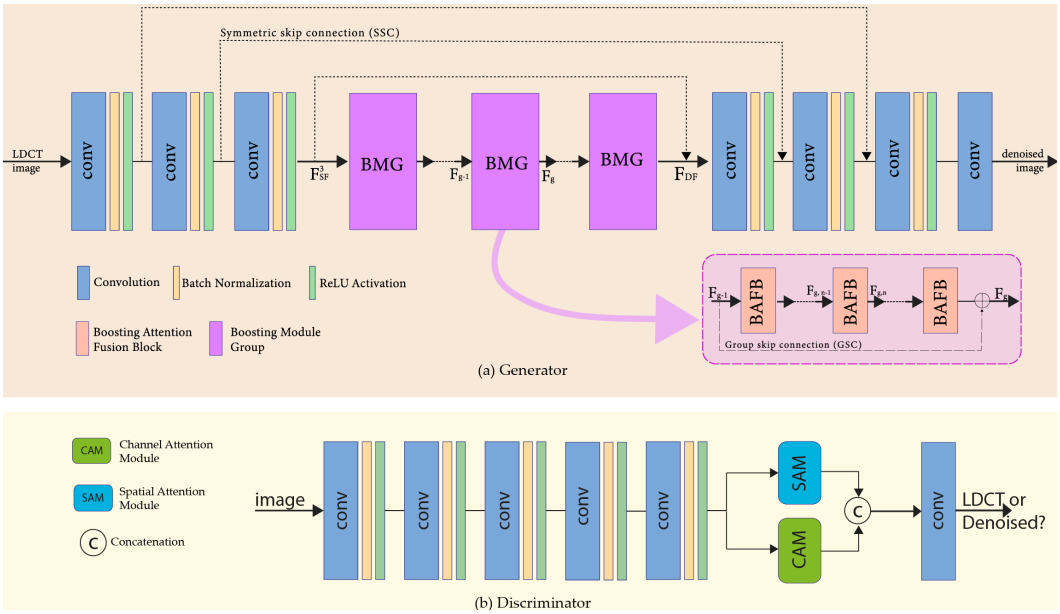


Fig. 1. The (a) generator (denoiser)and (b) discriminator network structure with boosting attention fusion blocks.

II. Methods

A. Loss Functions

The proposed model is composed of three main parts: (i) denoiser; (ii) discriminator and; (iii) feature extractor. Lyu et al. [3] first introduced the idea of Boosting Attention Fusion generative adversarial network (BAF-GAN). Their generator serves as the denoiser – which maps noisy image to a noise free one. The discriminator follows the basic discriminator system which is giving scores for the "candidate" image. One problem mentioned with the model was the iteration control during the boosting process. Further, they also had a difficulty with the stability and performance of their network. They have used pixel loss and structural similarity loss as loss functions in the discriminator and used VGG-19 for feature extraction. In this study, we proposed a combination of proper loss functions which is perceptual loss via the VGG16 pre-trained network [8] and MSE to be integrated in their model.

Mean squared Error (MSE) is one of the most common accuracy measurements that calculates the difference between the LDCT and ground-truth image pairs $(x_i, y_i)_{i=1}^{N}$ at a pixel level, and it is considered as one of the "per-pixel loss functions". This sums all the absolute errors between the pixels. Mathematically,

$$L_{MSE} = \frac{1}{N}\sum_{i=1}^{N}\|y_i - x_i\|^2 \tag{1}$$

Based from the previous studies, models like CycleGAN [12] and FFDNET [13] produced an over-smoothing problem along the edges of the CT images when using only MSE during the training process. In order to address this issue, the proposed model will be using the combination of MSE and perceptual loss, which can be calculated using the VGG16-pretrained network [8]. Unlike MSE, perceptual loss take high level features into consideration, which accurately models the human visual system due to its capability of learning the features. Similar to Ansari's DRL-network [5] and Ataei's cascaded CNN [4], the feature maps, ϕ_i, would be extracted from the last convolutional layer in blocks $i = 1, 2, 3, 4$ of the VGG-16 with size $h_i \times w_i \times d_i$ which can be expressed as:

$$L_{PL} = \sum_{i=1}^{N}\frac{1}{h_i w_i d_i}\left\|\phi_i(x) - \phi_i(y)\right\|^2 \tag{2}$$

B. Denoising Model

A typical GAN model is built from a minmax operation between the generator, G, in which the parameters map the samples (z) from the noise distribution $p(z)$, and the discriminator, D, which shows the probability that sample (x) belongs to true data $p_{data}(x)$. This structure can be represented as follows:

$$\min_G \max_D GAN(D,G) = E_{x\ p_{data}(x)}[\log D(x)] + E_{z\ p_z(z)}[\log(1 - D(G(z))] \tag{3}$$

Shown in Figure 1a is the full structure of the generator, G, acting as the denoiser. In this process, the LDCT image would have to go through pre-convolutional layer which composed of convolution, batch normalization (BN) and ReLU layers in order to extract the shallow features, F_{SF}^3. Next, multi-dimensional deep features, F_{DF}, would then be generated through the series of boosting module groups (BMG). For each BMG, a stack of n ϵ {1,..., N} boosting attention fusion blocks (BAFB), in which the spatial attention modules (SAM) and channel attention module (CAM) are being implemented. Finally, for the reconstruction layer,

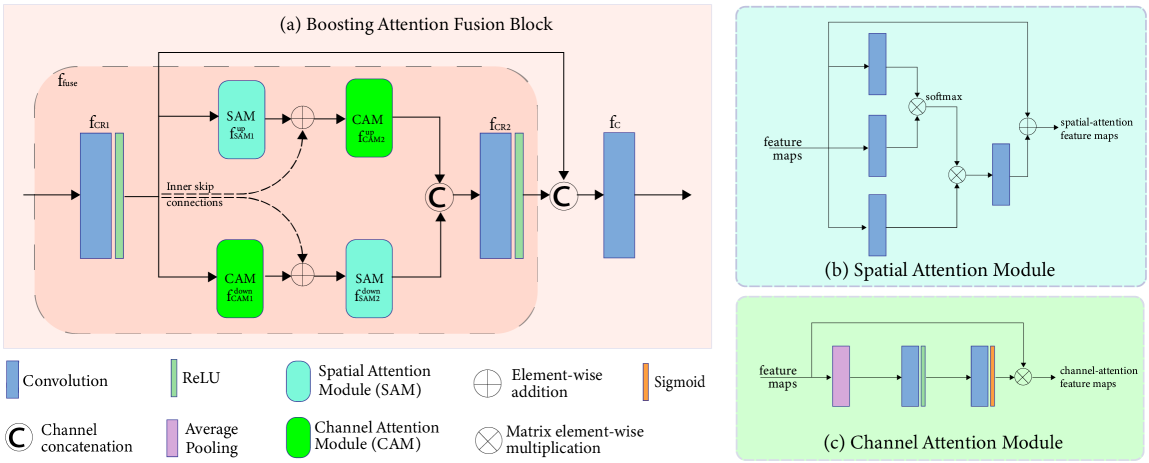


Fig. 2. The (a) boosting attention fusion architecture, (b) spatial attention module and (c) channel attention module.

$R_{postConv}$, deconvolution + BN + ReLU are applied. Overall, the denoiser architecture can be expressed as:

$$x_{denoised} = G(x_{LDCT}) \quad (4)$$

$$= G_{postConv}(G_{BMGs}(G_{preConv}(x_{LDCT})) \quad (5)$$

where $x_{denoised}$ is the output of the entire framework while x_{LDCT} is the input. Moreover, symmetric skip connections (SSC) were also applied to address the problem of vanishing gradient that is common in deep learning structures. This was done between the pre- and post- convolutional blocks. As for the discriminator, D, six convolution + BN + ReLU stacked layers were applied along with spatial- and channel- attention modules being integrated simultaneously to have more accurate detection of the images especially with discriminating channel-wise features. The structure for the discrimnator is shown in Figure 1b. Structural Similarity Index (SSIM) would be used for the comparison of the structural information of the images as in [5].

C. Spatial and Channel Attention Modules

Inside BAFBs from Figure 1, the integration of spatial- and channel- attention modules are implemented as shown in Figure 2a. This is due to the fact that simple Conv + BN + ReLU operation cannot capture the high and low frequency information of the feature map present during the pre-convolutional process. SAM ($f_{SAM}(\cdot)$), Figure 2b, is responsible for the long-range dependencies while CAM ($f_{CAM}(\cdot)$), Figure 2c, would capture the channel-wise features. However, these operations tend to bypass each other and therefore, fusion ($f_{fused}(\cdot)$) between the two should be applied. By implementing inner skip connections, the fusion block can be depicted as the following:

$$F_{up} = f^{up}_{SAM1}(f_{CR1}(x)) \oplus f_{CR1}(x) \quad (6)$$

$$F_{down} = f^{down}_{CAM1}(f_{CR1}(x)) \oplus f_{CR1}(x) \quad (7)$$

$$f_{fuse} = F_{up} © F_{down} © f^{up}_{CAM2}(F_{up}) © f^{down}_{SAM2}(F_{down}) \quad (8)$$

where $\oplus$ denotes element-wise addition and © represents channel concatenation in this case.

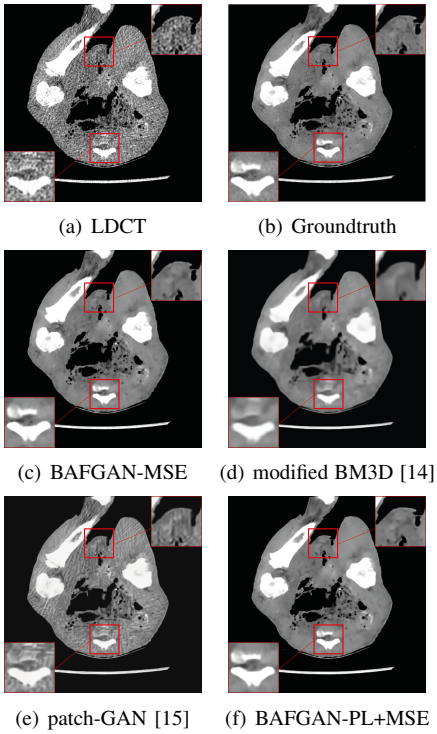
D. Data Preparation and Training Details

Most denoising models require a large dataset of normal-dose CT (NDCT) and LDCT image pairs, which are not readily available. This has been one of the difficulties in applying deep learning to denoising medical images. For this project, we have simulated LDCT from NDCT images following the simulation process in [5]. In this experiment, CT images were contained in DICOM files which include metadata like the image pixel size and medical information. One important thing to note when handling DICOM files is the measure of the radiodensity present in the CT images or also known as the Hounsfield unit (HU). CT images must be properly calibrated with the standard HU values first which corresponds to the specific subtance being observed in the images. For this experiment, the method for finding HU value follows the callibration method in [5].

A piglet dataset, obtained from a deceased piglet containing 900 slices taken with 100KVp, 0.625mm slice thickness, and using 300mAs for normal dose and 15mAs for low dose images. The dataset was divided into training set (70%) and testing dataset (30%). The training images of size 512 $\times$ 512 are subdivided into 40 $\times$ 40 overlapping patches in order to reduce the computational load of the network and increase the number of training samples. The training operation of the model maintain the same parameters in [3]. Both the generator and discriminator used Adam optimizer with learning rate of 0.0002, $\beta_1 = 0.01$, and $\beta_2 = 0.999$. The model was trained for 200 epochs with batch size of 4. Moreover, the implementation of this model was done with Tensorflow-Keras API on Windows operating system with Intel(R) Core(TM)i7 cpu @2.80 GHz processor and NVIDIA GeForce GTX 1080.

III. Experiment and Results

In order to verify the validity of the proposed structure of the loss function, three modifications of the model have been done: BAF-GAN with only MSE, BAF-GAN with only perceptual loss and BAF-GAN with the combination of the two loss functions as proposed. Further, the quantitative



(a) LDCT (b) Groundtruth

(c) BAFGAN-MSE (d) modified BM3D [14]

(e) patch-GAN [15] (f) BAFGAN-PL+MSE

Fig. 3. (a) Sample LDCT image from Piglet dataset along with the ground-truth image (b). The denoised image results from the models: (C) BAF-GAN with MSE, (d) BM3D, (e) modifies patch-GAN and (f) BAF-GAN with combination of MSE and perceptual loss.

results of the variation of the models were obtained by gathering the PSNR and SSIM of the models as summarized in Table I. The results were also compared with the results using DRLPS [4], BM3D [14] and self-attentive spectral normalized Markovian patch-GAN or modified patch-GAN [15] models. Figure 3 illustrates the visual results of the denoised images.

TABLE I
THE AVERAGE PSNR AND SSIM OF PIGLET DATASET

Models	**Piglet**	
	PSNR	**SSIM**
modified BM3D [14]	24.37	0.4461
modified patch-GAN [15]	30.37	0.5435
DRLPS [4]	32.18	0.5700
BAF-GAN-MSE	29.92	0.4987
BAF-GAN-Perceptual Loss	30.84	0.5888
BAF-GAN-MSE + Perceptual Loss	**33.64**	**0.7382**

Based from the results, the evaluation of the model with only MSE as the loss function has displayed over-smoothing along the edges and slight blurriness as observed in Figure 3(c). Looking at Figure 3(f), adding perceptual loss actually improved the quality of the denoised image due to the consideration of the differences between the images in various spaces and dimensions. Based on Table I, the PSNR and SSIM scores of the proposed model were slightly higher than the scores obtained by the other three models. The PSNR and SSIM of the modified patch-GAN are close to the values obtained by the model with only one loss function (either MSE or perceptual loss). However, there is still an apparent noise on the result gathered from the modified patch-GAN shown in Figure3(e), making the proposed model better visually.

IV. CONCLUSION

In this experiment, we show that creating feature maps by implementing the fusion of spatial- and channel- attention modules can enhance the signal-to-noise ratio of the images. Moreover, the effectiveness of perceptual loss in preserving structural details for denoising LDCT images was also observed. Finally, by taking advantage of the GPU's parallel architecture via GAN, the model becomes stable during the training process unlike the traditional iterative reconstruction LDCT denoising methods.

REFERENCES

[1] M.Li, W. Hsu, X. Xie, J. Cong, and W. Gao "SACNN: Self-Attention Convolutional Neural Network for Low-Dose CT Denosing with Self-Supervised Perceptual Loss Network," in *IEEE Transactions on Medical Imaging*, 2020

[2] D.Wu, H. Ren and Q.Li "Self-Supervised Dynamic CT Perfusion Image Denoising with Deep neural Networks," in *IEEE Transactions on Radiation and Plasma Medical Sciences*, pp. 1-1, 2020, doi: 10.1109/TRPMS.2020.2996566

[3] Q. Lyu, M. Guo, and M. Ma, "Boosting attention fusion generative adversarial network for image denoising," Neural Comput. Appl., vol. 0123456789, 2020, doi: 10.1007/s00521-020-05284-w.

[4] S. Ataei, J. Alirezaie, and P. Babyn, "Cascaded Convolutional Neural Networks with Perceptual Loss for Low Dose CT Denoising", in *2020 International Joint Conference on Neural Networks (IJCNN)*, doi: 10.1109/IJCNN48605.2020.9206816

[5] M. Gholizadeh-Ansari, J. Alirezaie, and P. Babyn, "Deep Learning for Low-Dose CT Denoising Using Perceptual Loss and Edge Detection Layer," J. Digit. Imaging, 2019, doi: 10.1007/s10278-019-00274-4

[6] K. Choi, J.S. Lim, and S. Kim "StatNet: Statistical Image Restoration for low-Dose CT using Deep Learning," in *IEEE journal of Selected Topics in Signal Processing*, 2020, doi: 10,1109/jstsp.2020.2998413

[7] A. Krizhevsky, I. Sutskever, and G.E. Hinton "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, Ed., vol. 25, Curran Associates, Inc. , 2012, pp.1097 - 1105.

[8] K. Simonyan, and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", 2014, arXiv: 1409.1556

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition", in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[10] H.S. Park, J. Baek, S.K. You, J.K. Choi, and J.K. Seo, "Unpaired Image Denoising Using A Generative Adversarial Network in X-Ray CT." in *IEEE Access*, vol.78, pp. 110414-110425, 2019

[11] Z. Yin, K. Xia, Z. He, J. Zhang, S. Wang, and B. Zu, "Unpaired Image Denoising via Wasserstein GAN in Low-Dose CT Image with Multi-Perceptual Loss and Fidelity Loss," in *Symmetry*, vol. 13, no. 126, doi:10.3390/sym13010126

[12] J. Gu and J. C. Ye, "AdaIN-Switchable CycleGAN for Efficient Unsupervised Low-Dose CT Denoising," pp. 1–12, 2020, [Online]. Available: http://arxiv.org/abs/2008.05753.

[13] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-Based image denoising," in *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, 2018, doi: 10.1109/TIP.2018.2839891.

[14] Y. Mäkinen, L. Azzari and A. Foi, "Collaborative Filtering of Correlated Noise: Exact Transform-Domain Variance for Improved Shrinkage and Patch Matching," in *IEEE Transactions on Image Processing*, vol.29, pp.8339-8354, 2020, doi: 10.1109/TIP.2020.3014721.

[15] S.Bera and P.K. Biswas, "Noise Conscious Training of Non-Local Neural Network powered by Self-Attentive Spectral Normalized Markovian Patch GAN for Low Dose CT Denoising," pp.1-12, 2020, [Online]. Available: https://arxiv.org/abs/2011.05684.