2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI 2018)

Motion-Blind Blur Removal for CT Images with Wasserstein Generative Adversarial Networks

Yilin Lyu[1], Wei Jiang[2], Yanjun Lin[1], Laszlo Voros[3], Miao Zhang[3], Boris Mueller[4], Borys Mychalczak[4], and Yulin Song[3], *Member, IEEE*

[1]Master of Arts in Statistics, Department of Statistics, Columbia University, New York, USA
[2]Department of Radiation Oncology, Yantai Yuhuangding Hospital, Yantai, China
[3]Department of Medical Physics, [4]Department of Radiation Oncology
Memorial Sloan Kettering Cancer Center, New York, USA

***Abstract* —Advanced deblurring techniques for computed tomography (CT) images are necessary and crucial to the improvement of accuracy of patient diagnosis in radiology and patient setup and treatment response assessment in radiation oncology. Currently, medical image deblurring is a challenging technical problem due to the unpredictability of patient motion. This paper introduces a new method of computed tomography image deblurring based on Conditional Generative Adversarial Networks (CGAN) that have been broadly implemented in computer vision research. A Wasserstein Generative Adversarial Network (WGAN) with adversarial loss and ℓ1 perceptual loss was proposed and trained by a blur-sharp image pair dataset created in-house and evaluated by Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM). These experiments showed the effectiveness of the approach, which outperforms other competing deblurring techniques both quantitatively and qualitatively.**

1. Introduction

The computed tomography (CT) scan is the most important imaging modality used for clinical staging of cancer. CT imaging is also integral to modern radiation oncology, especially in treatment planning, pre-treatment patient setup, and treatment response assessment. A significant challenge encountered in patients undergoing radiation therapy today is body and organ motion. Such motion can cause significant blurring on CT imaging that can adversely affect radiation planning and radiation treatment delivery.

The cone-beam CT (CBCT) installed on the modern medical linear accelerators requires more than 60 seconds for a full scan. Therefore, CBCT scans often exhibit pronounced motion-induced blur and artifacts. CT image deblurring is challenging because irregular patient body and organ motion can generate unpredictable blurring patterns. As a result, handcrafted programming is a suboptimal technique for removing blur. Creating an automated single-image deblurring method is therefore an important goal for radiation oncology physicists.

Yulin Song, Ph.D. is in the Department of Medical Physics, Memorial Sloan Kettering Cancer Center, Montvale, NJ 07645, USA (Corresponding author, phone: 201-775-7159, e-mail: songy@mskcc.org).

In recent years, there have been impressive achievements of Convolutional Neural Networks (CNN) that have revolutionized the field of computer vision. Generative Adversarial Networks (GAN) [1] are known for their ability to preserve high texture details in images and create solutions that are close to real image manifold and appear perceptually convincing. Meanwhile, CT image deblurring could be treated as a special case of such image-to-image translation.

Mathematically, a model can be trained to learn the mapping from blurry CT to sharp CT. The idea of DeblurGAN proposed by Kupyn and Budzan in 2017 [2] is innovative and creative. It includes two main parts, a generator (G) and a discriminator (D). G is trained to learn the mapping and to restore blurry CT to deblurred CT. Meanwhile, D is trained to gauge the similarity of the deblurred CT and the sharp CT. They are trained sequentially so that both G and D become synergistically stronger and ultimately the G can effectively remove the blur from blurry images.

The DeblurGAN is an approach based on Conditional GAN and a multi-component loss function [3]. Using Wasserstein GAN [4] with Gradient Penalty (WGAN-GP) [5] and perceptual loss, the DeblurGAN can achieve solutions that are perceptually hard to distinguish from real sharp CT and allow for restoring finer texture details in contrast to using traditional mean squared error (MSE) or mean absolute error (MAE) as an optimization target.

In this work, while using similar architecture with DeblurGAN, the loss functions were modified to push the model to restore more precise general content, which is more important for patient diagnosis, radiotherapy treatment planning, pre-treatment patient setup, and radiation treatment response assessment. The model was trained from scratch with a dataset using the blur generation algorithm proposed by [2]. The original sharp CT images were partly from the Cancer Imaging Archive (TCIA) of open-access medical image repositories, which include CT scans of patients with lung cancer, brain tumor, and other common diseases [6].

Section 2 of this paper introduces the methodology of the proposed approach. Section 3 briefly explains the blur

978-1-5386-7604-2/18/$31.00 ©2018 IEEE

generation algorithm and the dataset that was used in the project. Section 4 delineates the implementations of the approach. Section 5 shows its evaluation both quantitatively and qualitatively. Section 6 provides the conclusion and discussion.

2. METHODOLOGY

The primary aim of this research is to reconstruct a sharp CT given only a blurry CT as the input. Note that no prior information about the blur was provided. To achieve this goal, a CNN was designed and trained as generator G. For each blurry CT, it estimated the corresponding sharp CT. In addition, during the training phase, a critic network was introduced and trained as discriminator D. D was expected to distinguish between the generated restored CT and the sharp CT. Both networks were trained in an adversarial manner. For optimization, two types of losses were applied to the G. The first was Wasserstein WGAN-GP [5] loss and the second was $\ell 1$ Perceptual loss. Details about the loss functions are discussed in Section 2.2. Overall, the model of Conditional Wasserstein GAN with Gradient Penalty and Perceptual loss based on pretrained VGG-19 activations was suitable for the blur removal task. Such architecture also provided good results on other image-to-image translation problems such as super resolution, colorization, inpainting, raindrop removal, etc. [7,8,9,10]. Figure 1 shows the general architecture of the proposed model.

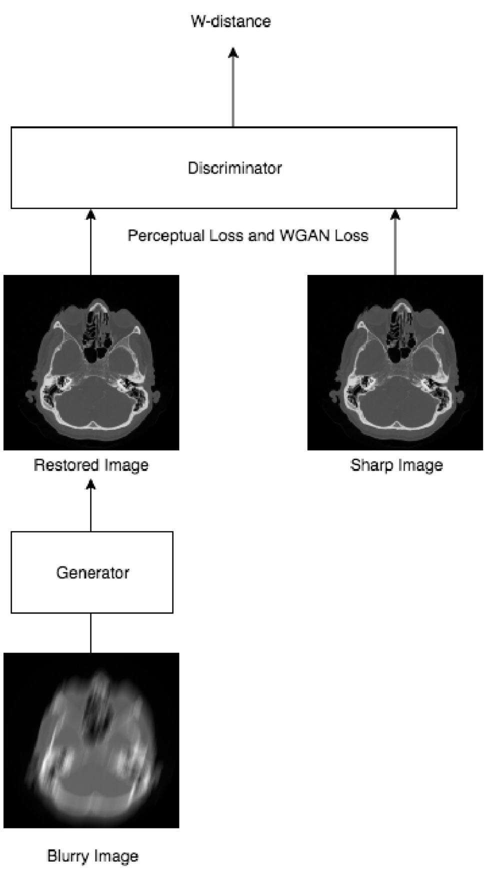


Figure 1. Work flow of Conditional Generative Adversarial Network for motion blur removal

2.1. Deep Learning Network Architecture

2.1.1 Generative Networks

The first two blocks are convolution blocks with a one-half stride. These are followed by nine residual blocks (ResBlocks) [11] and two transposed convolution blocks. Each ResBlock consists of a convolution layer, instance normalization layer, and ReLU activation. Dropout regularization [12] with a probability of 0.5 is added after the first convolution layer in each ResBlock. The generator also includes a global skip connection between the input and the last output of the transposed convolution block.

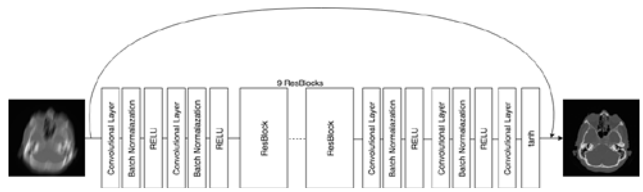


Figure 2. Architecture of the DeblurGAN generator network

1.1. Discriminative Networks

The Discriminator contains 4 convolution blocks. All convolutional layers contain batch normalization [13], except the first, and are followed by leaky ReLU activation.

2.1. Loss Functions

The loss function of the generator is formulated as a combination of adversarial loss L_{GAN} and the Perceptual loss $L_{Perceptual}$ [14]: $L = L_{GAN} + \lambda L_{Perceptual}$ The Perceptual loss, also known as the content loss, measures the global discrepancy between the features of the generator's output and those of the corresponding sharp images. These features can be extracted from a well-trained CNN. Here, a VGG19 trained on the ImageNet dataset is used.

The original DeblurGAN used a $\ell 2$ loss for Perceptual loss. According to [15], $\ell 1$ sparse regularization can cope with such changes much better than the standard $\ell 2$ norm. The results of experiments show that $\ell 1$ is more suitable for CT image deblurring, thereby preserving more details as expected. Thus, the Perceptual loss is:

$$E[(||VGG_l(I_s) - VGG_l(G(I_b))||_1]$$

For GAN, the goal of the generator and discriminator is to minmax the objective:

$$\min_G \max_D E_{x\sim P_r}\left[\log\left(D(x)\right)\right] + E_{\tilde{x}\sim P_g}\left[\log\left(1 - D(\tilde{x})\right)\right]$$

where P_r is the data distribution, and P_g is the model distribution implicitly defined by $\tilde{x} = G(z),\ z \sim P(z)$. The input z to the generator is sampled from the blurry CT. However, the divergences defined above that GAN typically minimizes are potentially not continuous with respect to the generator's parameters, leading to training difficulty. To solve this problem, Wasserstein GAN is proposed. In this method, the Wasserstein-I distance

$W\ (q,p)$ is used instead. Here, $W\ (q,p)$ is continuous everywhere and differentiable almost everywhere. The goal for both the generator and discriminator becomes:

$$\min_{G}\max_{D\in\mathcal{D}} E_{x\sim P_r}\left[D\ (x)\right] + E_{\tilde{x}\sim P_g}\left[D\ (\tilde{x})\right]$$

where $\mathcal{D}$ is the set of 1-Lipschitz function. To enforce the Lipschitz constraint on the discriminator, one needs to clip the weights of the discriminator to lie within a compact space $[-c,c]$. Clipping the weights in the discriminator would lead to optimization difficulties. Even when optimization succeeds, the resulting discriminator can have a pathological value surface. Therefore, Gradient Penalty is introduced as an alternative approach to enforce the Lipchitz constraint. A differentiable function is 1-Lipschitz if and only if it has gradients with norm of at most 1 everywhere, therefore directly constraining the gradient norm of discriminator's output with respect to its input. The new objective is:

$$E_{x\sim P_r}[D(x)] - E_{\tilde{x}\sim P_g}[D(\tilde{x})] + \alpha E_{\hat{x}\sim P_{\hat{x}}}[(||\nabla_{\hat{x}} D(\hat{x})||_2 - 1)^2]$$

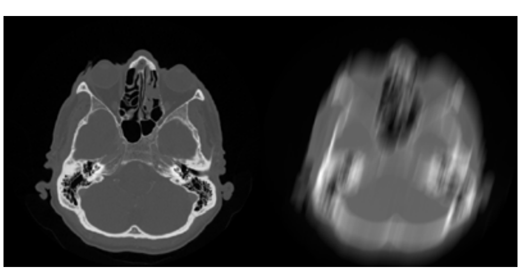
where $P\hat{x}$ is defined as sampling uniformly along straight lines between pairs of points sampled from the data distribution P_r and the generator distribution P_g. From experiments in [5], α =10 is used, which is believed to work well across a variety of architectures and datasets.

3. Paired CT Image Dataset

Similar to current deep learning methods, the GAN method described here requires a relatively large amount of data with ground truth for training. However, because there is no such public dataset for paired blur-sharp CT, one was created specifically for this project. Original sharp CTs were developed in part from open source datasets at TCIA [6]. Specifically, they were CPTAC-GBM, TCGA-SARC, CPTAC-LUAD, and CPTAC-HNSCC. CTs from different patients with different diseases are crucial to ensure the robustness of the model.

Registered free breathing (FB) CT scans and deep inspiration breath hold (DIBH) CT scans, provided by collaborators, were also combined to add reliable variance to the dataset. Reference [2] fully introduces a motion blur kernel generation algorithm. The randomness of this algorithm allows the models trained on a generated dataset to have enough generalization capacity.

This algorithm was used on the original dataset, and a new dataset was created with 14329 CT image pairs. Figure 3 shows examples of the generated blurry CT image and the corresponding sharp CT image.



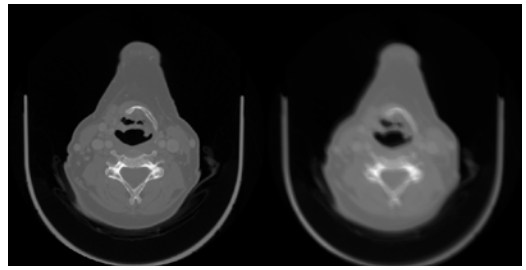


Figure 3. Examples from the paired dataset. Each pair, from left to right, is the sharp CT image and the blurry CT image. The first CT pair is generated with bigger blur parameters whereas the second pair is generated with smaller blur parameters.

4. Implementation

This method was implemented using the Tensorflow framework [16], and all the losses and their corresponding components were monitored using Tensorboard.

4.1. *Training Details*

The model was trained on a random crop of size 256 x 256 from the generated paired dataset consisting of 14329 pairs of blurry and sharp CT images acquired from various disease sites. For the parameters different from [2], adjustments were made, and greater weights were set for Perceptual loss to ensure that both losses were on a relatively similar scale, thereby showing empirically better results on validation. Because the models are fully convolutional and are trained on image patches, they can be applied to CT images with arbitrary sizes.

For optimization, first 5 gradient descent steps were performed on D, then one step was performed on G, using the Adam optimizer. The model was also trained by first training the G and then performing the previous recipe. After comparison, using the new recipe saved a lot of training time to achieve similar results by visual evaluation. The learning rate was set to 10^{-4} for both the generator and discriminator. Due to the high data volume, they were randomly split into 7 sub-datasets and were sequentially fed into the model. The batch size was set to 2 and the epoch numbers to 5 due to computation limitation. The training phase lasted roughly 40 hours per network. Figure 3 illustrates the losses of the method.

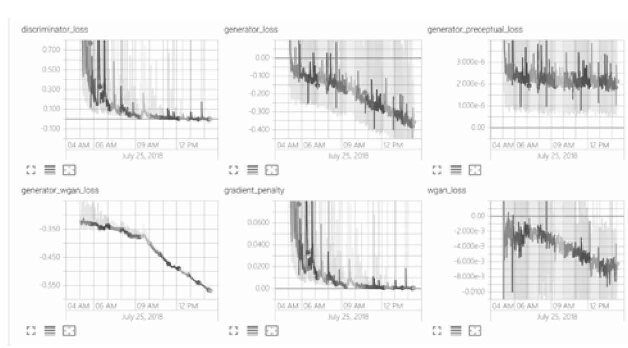


Figure 4. Tensorboard output of the training process. Obviously, all the losses showed a consistent trend of decreasing. The generator loss also had the potential to keep dropping; however, due to computational limitation,

the training process was terminated because the generated deblurred CT images were generally good enough for clinical application.

5. Results and Evaluations

5.1. *Quantitative Evaluation*

The whole dataset was randomly split into a training set and a testing set. The testing set, containing 300 images, was used for testing the robustness of the proposed model. This group of images was not used for training. Therefore, the model tried to deblur them for the first time. Both quantitative and qualitive evaluations are given below.

The model was tested on 300 CT images that were not used for training and compared with pre-trained DeblurGAN [2] on GoPro dataset and Pixel2Pixel [3] with the metrics of Peak Signal to Noise Ratio (PSNR) [2] and Structural Similarity Index (SSIM) [2]. Table 1 shows the evaluation results. Although not fully trained, the model still shows better results and great promise.

TABLE 1.
Mean Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM) measured on the test dataset of 300 images

	Our Method	**Pixel2Pixel**	**DeblurGAN**
SSIM	**0.94**	0.88	0.93
PSNR	**25.67**	24.72	25.22

5.2. *Qualitative Evaluation*

Because the method described here is flexible for input data sizes, a set of CT images with the original size were generated for visual evaluation by clinicians at MSKCC. The feedback was positive. Figures 5, 6, and 7 show representative examples of the generated deblurred CT images and the blurry CT images. From left to right are the deblurred CT images and the blurry CT images, respectively.

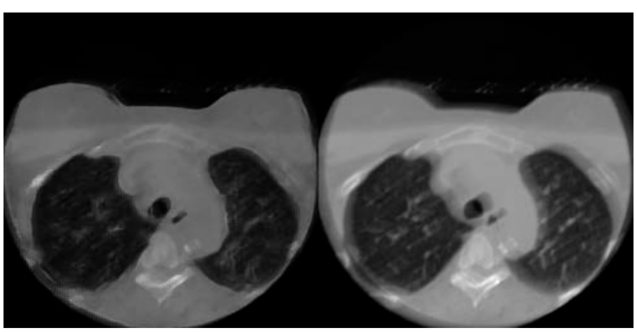


Figure 5. An example of a lung CT image. The lung-chest wall interface, trachea, pericardium, and body outline of the patient are clearly restored, showing great promise in accurate organ delineation in radiotherapy treatment planning.

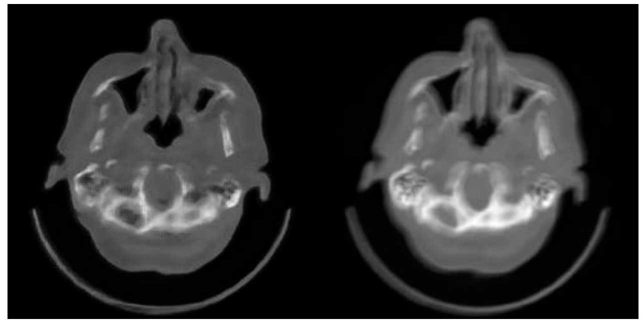


Figure 6. An example of a head CT image. The bony structure and para-nasal sinuses of the patient are obviously restored.

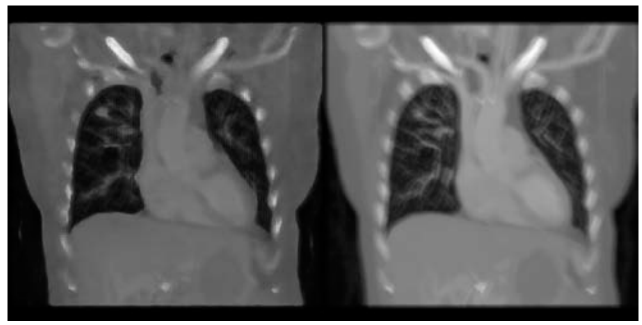


Figure 7. An example of a chest CT image.

6. Conclusions

In this research, the kernel-free blind motion deblurring approach was replicated to remove blur from CT images. A modified version of DeblurGAN [2] was implemented, which is a Conditional GAN optimized using a multi-component loss function. A paired dataset was created and the model was trained from scratch. This methodology has resulted in significantly improved performance on CT image blur removal.

In addition, the training procedure was optimized, which significantly reduced the training time without compromising clinical performance (quantitatively and qualitatively). Future studies include testing the sensitivity of the model on other medical image modalities to broaden its clinical applications.

Acknowledgement

This research was funded in part by the NIH/NCI Cancer Center Support Grant P30 CA008748. We would also like to thank TCIA for providing a great open-access source of CT dataset from various anatomical sites. The dataset collections from TCIA greatly augmented our dataset for training.

References

[1] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative Adversarial Networks. June 2014.

[2] O. Kupyn, et al. "DeblurGAN: Blind Motion Deblurring Using

Conditional Adversarial Networks." arXiv preprint arXiv:1711.07064, 2017.

[3] P. Isola, J.-Y. Zhu, T. Zhou, and A.A. Efros. “Image-to-image Translation with Conditional Adversarial Networks”. arXiv, 2016

[4] M. Arjovsky, S. Chintala, and L. Bottou. "Wasserstein GAN." arXiv preprint arXiv:1701.07875, 2017.

[5] I. Gulrajani, et al. "Improved Training of Wasserstein GANs." Advances in Neural Information Processing Systems. 2017.

[6] http://www.cancerimagingarchive.net/

[7] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. “Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network”. ArXiv e-prints, Sept. 2016

[8] O. Ronneberger, P. Fischer, and T. Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. arXiv e-prints, May 2015.

[9] R.A. Yeh, C. Chen, T. Lim, M. Hasegawa-Johnson, and M.N. Do. “Semantic Image Inpainting with Perceptual and Contextual Losses”. CoRR, abs/1607.07539, 2016.

[10] R. Qian, R.T. Tan, W. Yang, J. Su, and J. Liu, “Attentive Generative Adversarial Network for Raindrop Removal from a Single Image”. in Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, June 2018.

[11] K. He, X. Zhang, S. Ren, and J. Sun. “Deep Residual Learning for Image Recognition”. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.

[12] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”. Journal of Machine Learning Research, 15(1):1929–1958, 2014.

[13] S. Ioffe and C. Szegedy. “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift”. In International Conference on Machine Learning, pp. 448–456, 2015.

[14] J. Johnson, A. Alahi, and L. Fei-Fei. “Perceptual Losses for Real-Time Style Transfer and Super-Resolution”. In European Conference on Computer Vision, 2016. 1.

[15] Y. Chen, Y. Lai, and Y.J. Liu. CartoonGAN: Generative Adversarial Networks for Photo Cartoonization. Presented at: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Lake Salt City, USA, 18-22 Jun 2018.

[16] https://www.tensorflow.org/