applied sciences

MDPI

Supplementary Material

Reconstruction of Motion Images from Single Two-Dimensional Motion-Blurred Computed Tomographic Image of Aortic Valves Using In Silico Deep Learning: Proof of Concept

Yawu Long, Ichiro Sakuma and Naoki Tomii *

School of Engineering, University of Tokyo, Tokyo 113-0033, Japan
* Correspondence: tomii@g.ecc.u-tokyo.ac.jp

Overview

In this supplementary material, we give a detailed comparison analysis with 4 different architectures in inferring 60 motion images from a single motion-blurred CT image. These 4 different architectures are 2D U-Net [1], Pix2Vox [2], GAN [3] and the proposed model in the paper. Although these architectures are not designed for inferring 60 motion images from a single motion-blurred CT image, they also have the ability to solve ill-posed problems. 2D U-Net is a convolutional neural network that was developed for biomedical image segmentation problems. Motion images were laid on the channel dimension to enable the use of 2D U-Net. Pix2Vox is a framework for single-view 3D reconstruction. It is developed to infer 3D volume data from 2D images. GAN is a machine learning model in which two neural networks compete with each other to become more accurate in their predictions. These two neural networks are the generator and discriminator. We used the proposed model as a generator and a simple CNN as a discriminator (Figure S1). The loss function of discriminator is defined as the following:

$$L_{BCELoss} = -\frac{1}{n}\sum_i \left(t_i \cdot \log(o_i) + (1 - t_i) \cdot \log(1 - o_i)\right) \quad \text{(S1)}$$

where, the targets t_i should be numbers between 0 and 1, o is the label of the motion images. The loss function of generator is defined as the following:

$$L_g = L_{mse} + \lambda \cdot L_{BCELoss} \quad \text{(S2)}$$

where, L_{mse} is the mean square loss function. λ is weight parameters. We tried several values of λ, the value of 10^{-2} shows relatively higher accuracy.

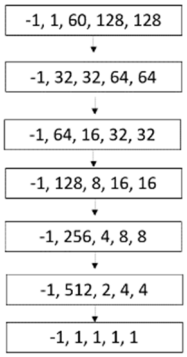

Figure S1. The architecture of the discriminator.

We trained all these architectures with the same training conditions mentioned in the paper. The running time for all models is shown in Table S1. Because an early stopping mechanism (the loss function doesn't decrease within 20 epochs) has been used, these models satisfied the condition at different epochs. The evaluation metrics employed include SSIM and

PSNR. The quantitative comparison can be found in Table S2 and Table S3. The performance of each architecture can be found in Figure S2. As observed, although there are differences in the performance of these architectures, the results show that DNN methods hold the possibility to infer motion images from motion blur in CT images. Even in the results of 2D U-Net and Pix2vox, a state of aortic valves from close to open can be expressed. And the reason that the performance of GAN is worse than the proposed architecture is the convergence failure in training GAN. The proposed model achieves good results on both SSIM and PSNR, especially that the images in rapid opening phase have the highest accuracy. In the evaluation of motions of aortic valves, motion images in the rapid opening phase are important to reflect the detailed opening process. Therefore, the results of the proposed architecture are relatively more valuable. Because the purpose of this paper is to prove the proposed concept, we only used the proposed model in the paper. And the superior accuracy in motion images of the rapid opening phase enables better evaluation of the performance in motion features.

Table S1. The running time for different architectures.

Architecture	Running Time
2D U-Net	38.5 h
Pix2Vox	63.0 h
GAN	70.8 h
Proposed model	76.2 h

Table S2. SSIM of different architectures.

	All images (N = 48000)	Images in the closed phase (N = 15983)	Images in the rapid opening phase (N = 16000)	Images in the slow closing phase (N = 16017)
2D U-Net	0.971±0.007	0.976±0.004	0.968±0.008	0.973±0.005
Pix2Vox	0.949±0.009	0.953±0.008	0.945±0.009	0.950±0.008
GAN	0.971±0.007	0.974±0.004	0.968±0.009	0.971±0.005
Proposed model	0.972±0.007	0.975±0.005	0.969±0.008	0.972±0.005

Table S3. PSNR of different architectures.

	All images (N = 48000)	Images in the closed phase (N = 15983)	Images in the rapid opening phase (N = 16000)	Images in the slow closing phase (N = 16017)
2D U-Net	35.964 ± 1.350	36.585 ± 1.132	35.496± 1.364	36.125 ± 1.109
Pix2Vox	30.195 ± 1.237	30.357 ± 1.234	30.049 ± 1.233	30.179 ± 1.224
GAN	35.390 ± 1.338	35.902 ± 1.013	34.877 ± 1.585	35.393 ± 1.142
Proposed model	35.962 ± 1.296	36.438 ± 1.110	35.523 ± 1.457	35.927 ± 1.123

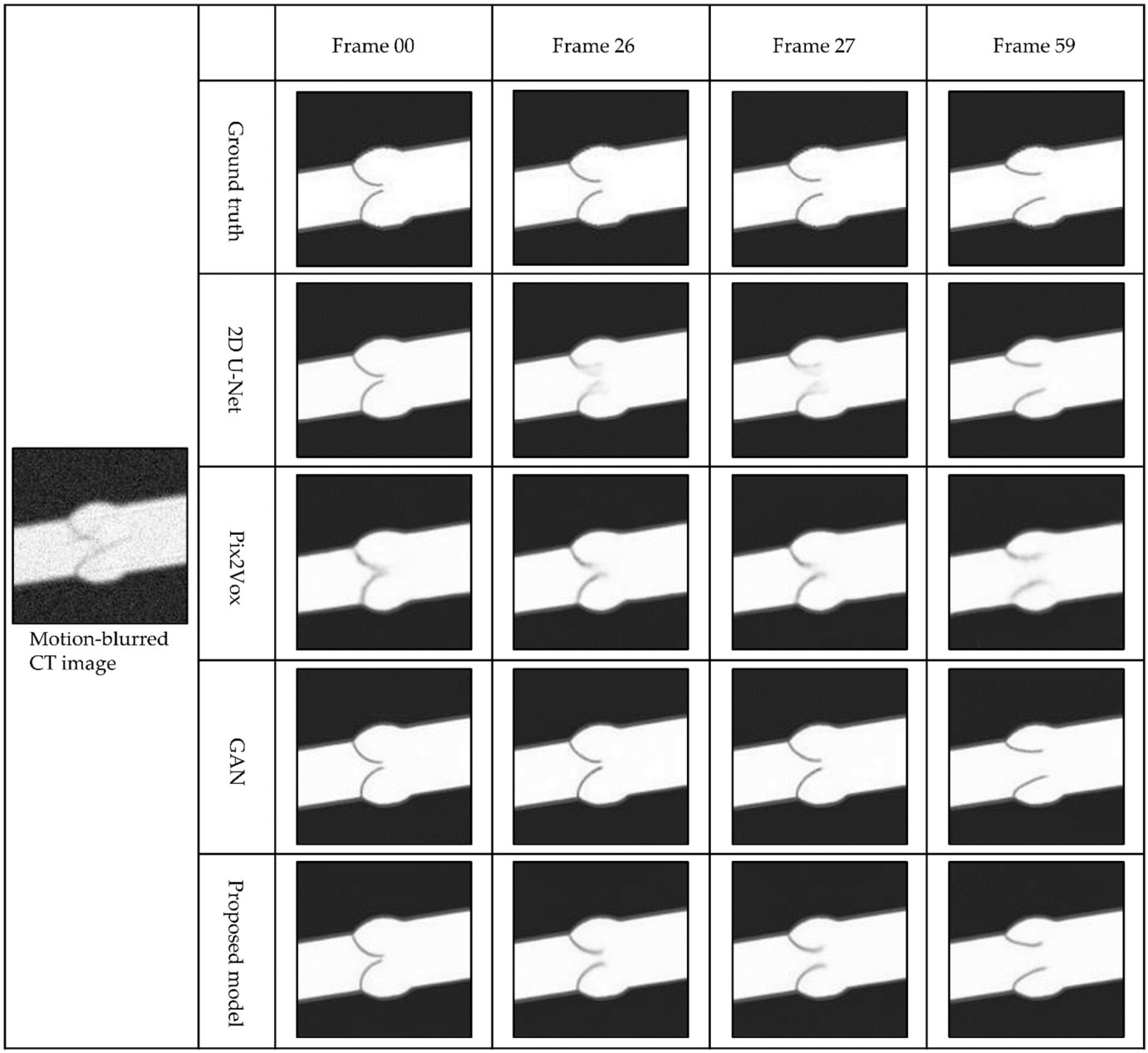

Figure S2. The performance of different architectures.

References

1. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention, Munich, Germany, 5-9 October 2015.
2. Xie, H.; Yao, H.; Sun, X.; Zhou, S.; Zhang, S. Pix2vox: Context-aware 3d reconstruction from single and multi-view images. In Proceedings of the IEEE/CVF international conference on computer vision, Seoul, Korea, 27th October to 2nd November 2019.
3. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courvillle, A.; Bengio, Y. Generative adversarial nets. Advances in neural information processing systems **2014**, *27*, 1-9.