A Cascaded Deep Learning Framework for Detecting Aortic Dissection Using Non-contrast Enhanced Computed Tomography

Xiangyu Xiong¹, Xiuhong Guan¹, Chuanqi Sun¹, Tianjing Zhang², Hao Chen¹, Yan Ding³, Zhangbo Cheng⁴, Lei Zhao³, Xiaohai Ma³ and Guoxi Xie¹

Abstract—Aortic dissection (AD) is a rare but potentially fatal disease with high mortality. The aim of this study is to synthesize contrast enhanced computed tomography (CE-CT) images from non-contrast CT (NCE-CT) images for detecting aortic dissection. In this paper, a cascaded deep learning framework containing a 3D segmentation network and a synthetic network was proposed and evaluated. A 3D segmentation network was firstly used to segment aorta from NCE-CT images and CE-CT images. A conditional generative adversarial network (CGAN) was subsequently employed to map the NCE-CT images to the CE-CT images non-linearly for the region of aorta. The results of the experiment suggest that the cascaded deep learning framework can be used for detecting the AD and outperforms CGAN alone.

Clinical relevance—This work provides a novel framework to synthesize CE-CT images from NCE-CT images and concludes a criterion to detect aortic dissection using the synthesized images.

I. INTRODUCTION

AD is a rare but life-threatening disease that has many complications [1]. It is associated with a high mortality of 1% per hour in untreated patients. Misdiagnosis of AD is devastating to the patients [2].

Accurate diagnosis of AD is an essential step for making correct treatment decisions to increase the patient’s chance of survival. Contrast enhanced computed tomography (CE-CT) is the first line technique widely used in detecting AD. As the X-ray decay coefficient of the contrast medium is much higher than that of the vessel wall, the aorta after injecting contrast medium becomes much brighter than vessel wall on CE-CT image. This makes the dissection symptom that the dark intima within the bright aorta was easily identified on CE-CT images. Thus, CE-CT has high sensitivity (>95%) and specificity (>95%) for the diagnosis of AD. However, CE-CT needs to inject contrast medium, which cannot be used in the patients with acute renal failure or allergic reactions [1], [3]. In addition, some AD patients have very similar symptoms to other cardiovascular diseases, these patients may receive non-contrast enhanced CT (NCE-CT) but CE-CT for disease examinations initially, leading to the misdiagnosis and missing the treatment time window for AD patients. If NCE-CT could provide hints for the diagnosis of AD patients, the misdiagnosis would be reduced. However, the dissection symptom on NCE-CT images is not obvious because the aorta without contrast medium has similar contrast with the intima.

In recent years, several models have been developed to synthesize a CE-CT image from a NCE-CT or low dose CE-CT image. These models are almost based on convolutional neutral network including generative adversarial network (GAN) [4]. In 2018, an encoder-decoder deep convolutional network was proposed to generate cardiac contrast enhanced CT images from contrast-free CT thoracic scans, with the purpose of volumetric assessment of left heart chambers [5]. Similarly, an encoder-decoder convolutional neural network was proposed to synthesize the full dose brain contrast enhanced MR images from the zero dose and low dose MR images to reduce gadolinium dose [6]. A steerable GAN method was proposed to generates absent MRA from existing MR multi-contrast images [7]. It is noteworthy that several studies have applied GANs in image synthesis of cross-modality for radiotherapy planning [8], [9], [10], [11], [12]. Inspired by the previous work, we proposed a novel framework to synthesize the CE-CT images from NCE-CT images to detect AD.

The key contributions of this work can be summarized as: 1. A novel deep learning framework combines a segmentation network with a conditional generative adversarial network (CGAN) to synthesize CE-CT images of high quality to detect AD. 2. Experiments are conducted to show that our proposed framework can synthesize better results compared with CGAN alone. 3. A reasonable criterion is concluded to demonstrate the improvements of detecting AD using the synthetic CE-CT images.
II. DATASET AND PREPROCESSING

A. Dataset collection

Total of 154 subjects, including 65 AD patients and 89 volunteers without AD but with other cardiovascular diseases, were recruited for collecting the paired datasets at two centers: Beijing Anzhen Hospital, Beijing, China, and Fujian Provincial Hospital, Fujian, China. All subjects volunteered to participate in this study which was approved by the Institutional Review Board. Each patient was sequentially conducted NCE-CT and then CE-CT with the same scan position, coverage, and parameters. The matrix size of acquired image is $512 \times 512$ with a resolution range from $0.625\text{mm} \times 0.625\text{mm}$ to $0.977\text{mm} \times 0.977\text{mm}$. The range of slice thickness is from 0.625 mm to 1.250 mm. The CT images were acquired from GE MEDICAL SYSTEMS, TOSHIBA, and SIEMENS devices with KVP from 100 to 120. ECG was used during data collection to reduce motion artifacts. Furthermore, the data was collected at the end of respiration with breath-holding to reduce misregistration of the two paired datasets.

B. Dataset selection

24 paired datasets from patients without AD were randomly selected for segmentation of aorta. The rest 130 paired datasets, including 65 AD patients’ datasets and 65 volunteers’ datasets, were further selected for synthetic network. Five-fold cross validation protocol was used to evaluate the performance of proposed framework, with a experimental setting (80% as training & 20% as testing).

C. Dataset Preprocessing

To further reduce misregistration of the two paired datasets obtain from each subject, we use elastix tool of Slicer3D software [13] to make a rigid registration for the paired NCE-CT and CE-CT images. To prepare the mask of aorta for training segmentation network, the aorta was then manually contoured from CE-CT images which were obtained from the 24 volunteers without AD via the Slicer3D software. The segmentation network was used to segment the aorta from NCE-CT images. The NCE-CT images and CE-CT images were then multiplied by the obtained masks of aorta to prepare a paired dataset for synthetic network.

It notes that NCE-CT and CE-CT images are resampled to the average voxel spacing of $0.731 \times 0.731 \times 0.951\text{mm}^3$ using linear interpolation. The paired volumes for synthetic network are cropped and padded to a size of $256 \times 256 \times 512$ for training acceleration. The volumes contain entire aorta that ranges from neck to pelvis. Due to abnormal CT value of aortic stent in the volume of postoperative patients, intensity values of NCE-CT volumes and CE-CT volumes is cut to $[0, 90]$ and $[0, 700]$, respectively.

III. METHODOLOGY

The cascaded deep learning framework was shown in Fig. 1, which contains a 3D segmentation network and a CGAN synthetic network. The 3D segmentation network was firstly used to segment aorta from the NCE-CT images and CE-CT images, and then CGAN was subsequently employed to map the NCE-CT images to the CE-CT images non-linearly for the entire aorta.

A. Segmentation of aorta

Encoder-decoder architectures based on deep convolutional neural networks, using skip connections to combine semantic information with spatial information, have been widely used in medical image segmentation, such as 2D U-Net [14] and 3D-UNet [15]. Compared with 2D U-Net, 3D U-Net has an advantage of using image information along with slice direction. In contrast to the networks which require expert knowledge and computing resources, nnU-Net [16] is recently developed as an out-of-the-box tool that can generate...
The segmentation experiment was conducted on a NVIDIA TITAN RTX GPU with 24GB memory. The patch size of 3D U-Net is defined as the size of $256 \times 256 \times 512$.

### B. Synthesis of CE-CT images

The synthetic network is CGAN, inspired by the GAN-based method proposed by Isola et al. [17]. As shown in Fig. 1(b), the generator of the CGAN is a “U-Net” like architecture that shares low semantic information between NCE-CT images and synthesized CE-CT images. The discriminator of the CGAN is based on a 70$\times$70 patchGAN [17]. It detects the synthesized CE-CT image is real or fake and guided the generator to synthesize images with sharp details.

To increase train samples for the CGAN, paired images are flipped by mirroring. We train the CGAN with an initial learning rate of $2 \times 10^{-4}$ via Adam optimizer, and the batch size is set as 1. The CGAN is implemented with Pytorch framework and we trained it on a NVIDIA TITAN GPU with 24GB memory for 200 epochs.

### IV. EXPERIMENTS AND RESULTS

#### A. Qualitative Evaluation

Our proposed framework is compared with CGAN with same train and test datasets. The CGAN synthesizes CE-CT images directly, whereas the proposed framework synthesizes CE-CT images for aorta only. As shown in Fig. 2, the proposed framework can synthesize clearer images than using CGAN alone. Specifically, in Fig. 2(a), (b), and (c), the proposed framework synthesizes clearer and smoother images for aorta without AD. In Fig. 2(d), the proposed framework synthesizes an intimal flap only in abdominal aorta (see the yellow box), whereas the CGAN synthesizes nothing in aorta (see the red box). In Fig. 2(e), the proposed framework synthesizes an intimal flap in thoracic aorta (see the yellow box), whereas the CGAN synthesizes a blurry and inhomogeneous aorta (see the red box). In Fig. 2(f), for a postoperative subject, the proposed framework synthesizes a winding intimal flap in abdominal aorta (see the yellow box), whereas the CGAN synthesizes two short and blurry intimal flaps near the aortic wall (see the red box).

#### B. Quantitative Evaluation

Three metrics, i.e., peak signal-to-noise ratio (PSNR), structural similarity (SSIM), mean absolute error (MAE), were used to evaluate the performance of CGAN and the proposed framework on synthesizing CE-CT images from NCE-CT images. For a fair comparison, all metric scores are calculated on pixels of aorta. Therefore, the region of aorta is extracted from the real CE-CT images and synthesized CE-CT images by multiplying the mask. As listed in Table I, the proposed framework significantly improves the PSNR scores and MAE scores and performs with a slight increase in SSIM scores compared to the CGAN.

#### C. Clinical diagnostic performance

Clinical performance of the proposed network was evaluated by an experienced radiologist. As the synthetic CE-CT images are essentially different from the real CE-CT images, the radiologist was trained to learn a criterion for the diagnosis of AD using a fold of the synthesized and corresponding real CE-CT first. The other four folds are then used to test clinical diagnostic performance. The criteria of determining AD on the synthetic CE-CT images includes a wider diameter of aorta, inhomogeneous CT density, a long intimal flap and consecutive short intimal flaps that are parallel or tilted to edge of aortic wall. Otherwise, the synthetic CE-CT images are defined without AD. Table II shows the diagnostic results using the synthetic CE-CT images.
V. DISCUSSION AND CONCLUSION

The aortic wall and blood have slightly different X-ray decay coefficients on NCE-CT images [18], making the proposed framework for synthesizing CE-CT images from NCE-CT images to detect AD become possible. It notes that the sensitivity of the proposed framework is high with an average of 96%. This indicates the proposed framework can reduce the misdiagnosis and prompt the patient to make a further examination. ECG simplifies the rigid registration and can lead to negligible registration errors. For future work, we will combine the synthesized network with an auxiliary convolutional neural network to automatically classify AD/Non-AD subjects using the synthesized CE-CT images.

In summary, our experimental results demonstrate that the proposed framework can be used for detecting AD and significantly outperformed CGAN alone.

Table I

<table>
<thead>
<tr>
<th>Experiments</th>
<th>CGAN</th>
<th>PSNR</th>
<th>SSIM</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fold1</td>
<td>31.15±3.049</td>
<td>0.986±0.004</td>
<td>2.367±1.470</td>
<td></td>
</tr>
<tr>
<td>Fold2</td>
<td>30.54±3.654</td>
<td>0.986±0.1006</td>
<td>2.499±1.329</td>
<td></td>
</tr>
<tr>
<td>Fold3</td>
<td>32.07±4.098</td>
<td>0.987±0.004</td>
<td>2.105±1.156</td>
<td></td>
</tr>
<tr>
<td>Fold4</td>
<td>31.602±3.451</td>
<td>0.985±0.006</td>
<td>2.143±0.957</td>
<td></td>
</tr>
<tr>
<td>Fold5</td>
<td>31.102±3.335</td>
<td>0.985±0.006</td>
<td>2.373±1.287</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>31.297±3.763</td>
<td>0.985±0.005</td>
<td>2.297±1.261</td>
<td></td>
</tr>
</tbody>
</table>

Table II

<table>
<thead>
<tr>
<th>Experiments</th>
<th>Proposed framework</th>
<th>PSNR</th>
<th>SSIM</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fold1</td>
<td>33.77±2.347</td>
<td>0.990±0.003</td>
<td>1.247±0.561</td>
<td></td>
</tr>
<tr>
<td>Fold2</td>
<td>32.639±3.691</td>
<td>0.989±0.005</td>
<td>1.654±1.278</td>
<td></td>
</tr>
<tr>
<td>Fold3</td>
<td>34.590±3.088</td>
<td>0.991±0.003</td>
<td>1.196±0.787</td>
<td></td>
</tr>
<tr>
<td>Fold4</td>
<td>33.806±3.446</td>
<td>0.990±0.004</td>
<td>1.375±0.917</td>
<td></td>
</tr>
<tr>
<td>Fold5</td>
<td>33.447±2.915</td>
<td>0.989±0.004</td>
<td>1.377±0.731</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>33.652±3.271</td>
<td>0.990±0.004</td>
<td>1.369±0.899</td>
<td></td>
</tr>
</tbody>
</table>

References