Applying Cross-modality Data Processing for Infarction Learning in Medical Internet of Things

Chenru Xu, Zhifan Gao, Dong Zhang, Jinglin Zhang, Lei Xu, and Shuo Li

Abstract—Cross-modality data processing is critical for the internet of things (IoT) deployment in healthcare. It can convert the innumerable raw day-to-day medical big data from massive IoT-based medical devices to diagnostic valuable data so that they can be fed to clinical routine. In this paper, we propose a novel spatiotemporal two-streams generative adversarial network (SpGAN) as a cross-modality data processing approach to deploy the medical IoT in infarction learning. Our SpGAN remotely converts diagnostic valuable contrast-enhanced images (the 'gold standard' for infarction learning, but it requires the injection of contrast agents) directly from raw non-enhanced cine MR images. This converting allows physicians to remotely perform infarction observation and analysis to break through the limitations of time and space by building a cloud computing platform of IoT-based MRI devices. Importantly, this converting offers a low-risk IoT-based manner to eliminate the potential fatal risk caused by contrast agent injection in the current infarction learning workflow. Specifically, SpGAN consists of 1) a spatiotemporal two-stream framework as an encoding-decoding model to achieve data converting, 2) a spatiotemporal pyramid network enhances those features which are responsible to the infarction learning during encoding to improve decoding performance. Real IoT-based remote diagnosis experiments performed on 230 patients demonstrate that SpGAN provides high-quality converted images for infarction learning and promotes the in-depth application and deployment of IoT in the medical field.

Index Terms—Internet of things, cross-modality feature learning, Two-streams framework, Contrast-enhanced imaging, Adversarial learning.

I. INTRODUCTION

Cross-modality data processing capability is critical for the internet of things (IoT) in healthcare [1], [2]. Recently, IoT systems have been merged into medical field to create the medical IoT for generating and collecting a big volume of multimodal medical data, i.e., medical big data. Exploring the best diagnostic decision or wisdom from the medical big data to feed into the healthcare system is becoming extremely desirable. Cross-modality data processing meets this desire and it can dig out cross-modality concealed symptom information to obtain valuable diagnostic data from the raw data generated by IoT systems [4], thereby providing corresponding treatment information to serve smart health care [5], [6]. With the advanced artificial intelligence technology, it is capable to enable the efficiency in significant health characteristics analysis, as well as the construction of inference to different scenes [7], [8]. It provides a high-performance foundation for remote diagnosis and significantly reduces the time, cost, and risk of medical IoT deployments [9].

Deploying a cross-modality data processing-based medical IoT application has more significance regarding the learning of infarction. This because it has great potential to eliminate the use of contrast agents (CAs) [10]. In the current clinic setting, the common raw data of IoT-based MRI is non-enhanced MR image data. This data is greatly limited in infarction diagnosis compared with the contrast-enhanced MR image data that is the ‘gold standard’ but needs to use CA [11]. However, contrast-enhanced MR image data and its CAs have been criticized by their potential toxicity, even the fatal risk for those patients with compromised kidney function. The cross-modality data processing-based medical IoT has the ability to convert the raw non-enhanced MR image data into the valuable diagnostic contrast-enhanced MR image data for radiologist diagnosis [12]. Therefore, it will provide remote infarction diagnoses to local patients who are limited in the use of contrast agents (CAs) [13], and eliminate CA-associated health risks, resulting in substantial improvements in current infarction clinic workflow [14].

There are few well-established cross-modality data processing approaches to convert the raw non-enhanced MR image data into the valuable diagnostic contrast-enhanced MR image data, except for some attempts. Some attempts [15], [16] are used to assist scar segmentation in infarction learning. They convert a single T1 MR image to a rough contrast-enhanced image. However, the information provided by a single T1MR image is limited, which causes the quality of this rough contrast-enhanced image is greatly far away from the real contrast-enhanced image. Thus the rough contrast-enhanced image has no ability to learn the infarction effectively. Another attempt [17] is used to expand the training data in the task of infarction scar segmentation. It obtains a random contrast-enhanced image from random noises. However, the random contrast-enhanced image generation cannot correspond to the correct infarction status of the patient and thus is useless in real clinic patient-based infarction learning.

Converting non-enhanced MR image data into contrast-enhanced MR image data has multiple challenges. 1) Recent
work shows that cine MR images may be the best choice for non-contrast image data during cross-modality infarction learning because it provides more comprehensive information than a single MR image. But the cine MR images are 2D+T data and the contrast-enhanced MR image is a 2D image. The complex and variable spatiotemporal representation within the 2D+T images often leads to its low efficiency to build an accurate nonlinear mapping to the 2D images during cross-modality learning [18], [19]. 2) the infarction has invisibility, i.e., the infarction is only visible in contrast-enhanced images after CA injection and is invisible in non-enhanced images without CA injection. But the surrounding background tissues (e.g., myocardium, papillary muscles, and blood vessels) are not significant changed. This leads to the infarction that may be interfered by surrounding background tissues during cross-modality learning [20], [21]. 3) Different patients require unique cross-modality learning that corresponds to the patients personal information. Different patients have large diversities in the location, shape, and size of various cardiac tissues (including infarction). Thus, finding a systematic learning model to understand all of these large diversities is difficulty[22], [23].

In this paper, we propose a novel cross-modality data processing named spatiotemporal two-streams generative adversarial network (SpGAN). The SpGAN deploys medical IoT in infarction learning by converting raw non-enhanced cine MR images into diagnostic valuable contrast-enhanced images (converted infarction images). Our novel SpGAN remotely encodes the different spatiotemporal motion pattern of each myocardial tissue (including the infarction) from the cine MR images, and decodes these spatiotemporal patterns to learn the infarction through the presentation of the contrast-enhanced images. Specifically, our SpGAN splits both a generator and a discriminator into two pathways: the generator initially uses a structural pathway to learn structural features between all tissues by using multi-range tissue structural dependencies. Simultaneously, the generator uses a detailed pathway to learn the detailed features of the scar tissue by coordinating local fine details. Then, SpGAN complementarily combines these two pathways to convert a contrast-enhanced infarction image. The discriminator uses a structural crit to assess the structural content of the converted image by classifying the authenticity of the image and uses a detailed crit to assess finely detailed features of the converted image by minimizing the output difference of each convolutional layer. Moreover, SpGAN embeds hidden personalized information into the generators and discriminators by an attention-driven self-learning process instead of a priori label. This information is a constraint for improving the extremely personalized tissue during the entire adversarial training. Furthermore, our method introduces a synergy loss function which combines two recent loss functions and a novel L2 overlapping group sparsity anisotropic total variation to improve the training effectiveness and stability.

Thus, the contribution of this work is divided into the following four items:

- For the first time, a cross-modality data processing approach is proposed to deploy medical IoT in the infarction learning. It will assist remote infarction diagnoses to patients who are limited in the use of (CA), and eliminate CA associated health risks, reduce cost in the workflow.
- A new spatiotemporal two-streams framework is proposed to synthesize personalized medical images. This framework effectively integrates global structural information to facilitate the accuracy of tissue physiological structure of converted images. It also integrates local detailed information to enhance the authenticity within each tissue detail of converted images.
- A novel spatiotemporal pyramid network is proposed to encode and decode spatiotemporal representation from the 2D+T images to the 2D images. This architecture is able to not only enrich the spatiotemporal representation of all cardiac tissues by the spatiotemporal feature pyramid, but also strengthen the task-specific representation in the direction of performance improvement by a spatiotemporal attention.
- A novel loss term is proposed to improve the stability and smoothness of the non-stochasticity GAN generation by the L2 overlapping group sparsity anisotropic total.
I. INTRODUCTION

Our novel SpGAN builds a cross-modality feature mapping $\mathcal{M}$ from the 2D+T cine MR images (non-enhanced data) $C$ to a 2D converted infarction image (contrast-enhanced data) $I_p$. As shown in Fig. 2, the network consists of two main modules. Firstly, a generator $G(C)$ is trained to convert an image $I_p$, from $C$ by two pathways (structural pathway $G_s$ and detailed pathway $G_d$). Importantly, $p_s$ is self-learned from $C$, which requires neither a priori personalized label nor the expected target image $S_p$ [24]. Then, the discriminator $D(I_p, p_s)$ evaluates the quality of the converted image by two pathways (structural critic $D_s$ and detail critic $D_d$). Both the generator and the discriminator effectively integrate multi-attention mechanism models with multi-range, hierarchical dependencies. Thus, they can understand both structural maps and fine details from all feature locations during converting, thereby improving the realism of the converted infarction image.

A. Network Architecture

1) Generator $G$: The generator $G$ is split into a structure pathway $G_s$ and a detailed pathway $G_d$. $G_s$ learns the global structural information of all tissues, and $G_d$ learns local detailed information about the scar tissue. These two pathways complement each other to improve the quality of the converted infarction image.

Structure pathway $G_s$ creates a spatiotemporal pyramid network (Sect. II-B1) to equally learn multi-range, hierarchical structural dependencies among all tissues (pixels) to present the structural image $I_s$, which focuses on a general structure and the image content. Concretely, this spatiotemporal feature pyramid employs the stack Bi-ConvLSTM block [25] to encode the input $C$ into a latent vector and then decode this vector into an image $I_s$. Thus, this pathway effectively ensures the patients physiological structural accuracy in the converted image.

Detailed pathway $G_d$ adopts a spatiotemporal attention network (Sect. II-B2) to unequally learn fine details by assigning different weights to different tissues (pixels) to present detailed image $I_d$, which focuses on the fine details of the scar tissue. This pathway takes the $C$ and enables the model to only focus on those spatiotemporal features of $C$ which are responsible for generating the infarction and keeping the remaining features of the background tissues. Then, the integrated feature of $p_s$ and $I_s$ are fed into a decoder network to output an image $I_d$.

Then, $G$ adopts a concatenation operation to construct the final converted image by combining the structural pathway output $I_s$ and the detailed pathway output $I_d$. The final image is described as follows:

$$I_p = (1 - \alpha) I_d + \alpha(1 - I_s) \odot I_s, \quad (1)$$

where $\alpha$ is a balance parameter and the $\odot$ denotes the element-wise multiplication operator. Thus, generator $G$ is defined as:

$$G(C) = (1 - \alpha)G_d(G_s(C), p_s) + \alpha(1 - G_d(G_s(C), p_s)) \odot G_s(C). \quad (2)$$

2) Discriminator $D$: The discriminator is split into a structural critic $D_s$ and a detailed critic $D_d$, where $D_s$ assesses image structural content and $D_d$ assesses texture details; therefore, a more comprehensive assessment of converted images is achieved through these two critics synchronous adversarial learning with the generator. $D$ firstly maps the input image to multiple overlapping patches (PatchGAN [24]) and then penalizes the image by the two critics convolutional outputs. $D$ also combines hidden personalized spatiotemporal information.
Fig. 3. Our method uses 1) $G_s$ and $G_d$ to focus on learning global structural dependencies and tissues by stacked Bi-ConvLSTM blocks and local fine details of tissues by a combination of non-local block and U-Net; 2) $D_s$ and $D_d$ to map the input image to multiple overlapping patches and penalize the image-focused structure content by the authenticity loss and the texture detail by the detailed loss.

$p_s$ to enhance the presentation of personalized synthesis during the adversarial learning.

**Structural critic** $D_s$ optimizes the structural content authenticity of the converted infarction image. $D_s$ averages all the patches responses to provide an overall classification result whether an image is real or fake. Inaccurate classifications of the image are penalized by the authenticity loss term.

**Detailed critic** $D_d$ optimizes the texture details authenticity of the converted infarction image. $D_d$ computes and discriminates the difference in the output of each convolution layer between the converted infarction image and the real contrast-enhanced image. The patches convolutional feature that mismatches at any layer in the image is penalized by the detail loss term.

By sharing the weights of $D_s$ and $D_d$, the discriminator simultaneously evaluates multiple-level features of the converted image while ensuring the training stability.

**B. Network Detail**

1) **spatiotemporal pyramid network**: spatiotemporal pyramid network is designed to accurately and comprehensively learn the spatiotemporal information on the morphology and kinematics of all myocardial tissues from the cine MR images. This network is modified from the network designed by Xu et al. [26]. Concretely, the spatiotemporal feature pyramid innovatively builds pyramidal features to obtain a feature hierarchy consisting of 3-scaled feature maps from the 2-4 ConvLSTM Blocks. The benefit of hierarchical pyramid features is two-folds: 1) by using multiple scale feature learning frameworks, the encoder handles the asymmetry distortion that is a mutable myocardial motion and deformation pattern; 2) a pyramid feature is used to aggregate features of different receptive fields in encoding and improve the spatial resolution in the decoding.

2) **Spatiotemporal attention network**: the spatiotemporal attention network is designed to selectively learn the fine and detailed spatiotemporal information on infarction. It combines an attention block [27] with a U-net architecture [28]. Concretely, this network firstly creates an attention-weighted feature by leveraging a $5 \times 5 \times 5$ 3D convolution layer with a normalization layer, a non-local network, and a fully connected layer. This feature is then concatenated with a $64 \times 64 \times 64$ feature from the last layer of $G_s$ and fed into the U-net. The benefit of attention-weighted features is two-folds: 1) by computing the response at a position as a weighted sum of the features at all positions, features are adaptively weighted according to their contributions to the task target; 2) by computing the correlation between any two positions in feature maps, regardless of their positional distance, the learning ability of the spatial relationship of image pixels is improved from short-range to long-range.

**C. Loss function**

Our SpGAN incorporates a synergy loss, which greatly improves adversarial learning by integrating three complementary loss terms: the **authenticity loss** promotes the distribution of the converted images to be closer to that of the ground truth; the **boot loss** drives the smoothness of instances in the structure image $I_s$ and prevents them from randomness; and the **detail loss** favors focusing on restoring texture details in the converted image $I_{p_s}$.

**Authenticity loss.** To maximize the probability that the converted infarction image fools the discriminator as the real CA-enhanced images, the authenticity loss is adopted from WGAN-GP [29]. Specifically, WGAN-GP avoids potential discontinuity with respect to the generators parameters and local saturation leading to gradient-vanishing by using a continuous earth mover distance to replace the Jensen-Shannon divergence distance in the GAN formulation. Moreover, WGAN-GP also adds a gradient penalty to penalize the norm of the gradient in the discriminator about its input, thereby avoiding undesired behaviors caused by enforcing a Lipschitz constraint on the discriminator.

Thus, the authenticity loss $\mathcal{L}_a(G, D_s, I_{p_s}, C)$ formula is as follows:

$$
\mathcal{L}_a(G, D_s, I_{p_s}, C) = \mathbb{E}_{I_{p_s} \sim P_g} [D_s(G(C))] - \mathbb{E}_{I_{p_s} \sim P_{data}} [D_s(I_{p_s})] + \lambda_1 \mathbb{E}_{I \sim P_x} [\|\nabla I D_s(\hat{I})\|^2 - 1]^2,
$$

(3)
The boot loss does not affect the detailed pathway and even the entire generator. The detailed loss evaluates the error distance of details between the converted image and the ground truth, the detailed loss is modified from the recently proposed perceptual loss [33]. The detailed loss is defined as:

\[ \mathcal{L}_d(G, D, \mathcal{P}_s, \mathcal{C}) = \lambda_1 \mathcal{L}_{\mathcal{P}}(G, D, \mathcal{P}_s, \mathcal{C}) + \lambda_2 \mathcal{L}_b(G, D, \mathcal{P}_s, \mathcal{C}), \]

where \( \lambda_1 \) and \( \lambda_2 \) are the hyperparameters that balance the relative importance of the different terms. Finally, we define the following mini-max problem:

\[ \arg \min_{G} \max_{D, D_a, D_b} \mathcal{L}. \]  

III. MATERIALS AND IMPLEMENTATION

A total of 230 infarction patients (Table I) with short-axis MR images and contrast-enhanced images are selected to train our method. This retrospective study was approved by our institutional review board in accordance with local ethics procedures. Further consent was waived with approval. MR imaging is performed using a 3T MR system. Infarction contrast-enhanced imaging is performed in the same orientations and with the same slice thickness as those of cine imaging ten minutes after the intravenous injection of gadolinium (0.2 mmol/kg). The generator changes the batch normalization to instance normalization. The discriminator adopts the PatchGAN architecture of [23] but removes feature normalization. It computes the gradient penalty with respect to each input independently [34]. The weights of the two discriminator pathways are shared. A network from [33] is used to automatically crop the cine MR and contrast-enhanced images to \( 64 \times 64 \) ROI sequences, including the LV. All networks are trained using an ADAM solver [36] with the batch size 8 and an initial learning rate of 0.001. For every 5 optimization steps of the discriminator, we perform a single optimization step of the generator. The hyperparameters in the network are set as: \( \alpha = 0.84, \lambda_1 = 10, \lambda_2 = 0.25, \) and \( \lambda_3 = 50. \) The pixel values are normalized to \([0, 1]\).

SpGAN connects a generator and a discriminator by taking the output of the generator as the input of the discriminator. Both the generator and the discriminator have a structural pathway and a detailed pathway, respectively.

<table>
<thead>
<tr>
<th>Gender(male)</th>
<th>177</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age(yr)</td>
<td>56.4 ± 10.1</td>
</tr>
<tr>
<td>Weight(kg)</td>
<td>75.3 ± 9.7</td>
</tr>
<tr>
<td>Height(cm)</td>
<td>167.1 ± 8.8</td>
</tr>
<tr>
<td>Left ventricular ejection fraction(%)</td>
<td>35.0 ± 18.1</td>
</tr>
<tr>
<td>Stroke volume(mL)</td>
<td>40.6 ± 17.2</td>
</tr>
<tr>
<td>Coronary risk factors</td>
<td>Hypertension 124</td>
</tr>
<tr>
<td>Smoking</td>
<td>119</td>
</tr>
<tr>
<td>Family history</td>
<td>13</td>
</tr>
</tbody>
</table>

TABLE I
DEMOGRAPHICS OF PATIENTS

Authorized licensed use limited to: UNIVERSITY OF WESTERN ONTARIO. Downloaded on March 26,2021 at 16:49:09 UTC from IEEE Xplore. Restrictions apply.
The generator inputs the 2D+T cine MR images $25 \times 64 \times 64 \times 1$, where $H = W = 64$ are the height and width of each temporal frame, $T = 25$ is a temporal step, $C = 1$ is the number of channels. This input then is fed to the structural pathway and detailed pathway of the generator to output a structural image $64 \times 64 \times 1$ and a detail image $64 \times 64 \times 1$. These two images are combined by Eq.1 to obtained the converted image $64 \times 64 \times 1$ that is the output of the generator.

The discriminator inputs the converted image $64 \times 64 \times 1$ and the hidden personalized spatiotemporal information $32 \times 32 \times 64$ extracted from cine MR images. These two input then are fed to the structural pathway and detailed pathway of the discriminator. The output of discriminator is the outputs of the two pathways, which are the real or fake -1, 1 based on Authenticity loss and the value of feature errors based on Detail loss. The discriminator with its two pathway are used only during adversarial training.

Our SpGAN is trained by 138 patients and is independently tested by 92 patients. The performance of the SpGAN has been evaluated by 1) three image similarity metrics between the converted infarction images and the clinical contrast-enhanced images; 2) three clinical metrics of consistency between these two types of images in IoT-based remote diagnosis. For image similarity metrics, the structural similarity index (SSIM), peak signal-to-noise ratio (PSNR), and normalized root-mean-square-error (NRMSE) between the converted infarction images and the contrast-enhanced images are calculated. For clinical consistency metrics, two radiologists (> 10 years in experience) remotely and manually diagnose the term of scar tissue localization (16-segment model), scar size, and scar tissue ratio (% scar size / myocardium) in both converted infarction images and contrast-enhanced images.

### IV. Experiments and Results

The experiment indicates that the proposed SpGAN obtains highly realistic contrast-enhanced images in real IoT-based infarction learning.

#### A. High consistency in IoT-based remote infarction learning

SpGAN yields a strong consistency when comparing its converted infarction images with the real contrast-enhanced images in the IoT-based remote infarction diagnosis. When the radiologists localized the infarction (16 segments) for these two images, 90.8% (209 of 230) of consistency between the two localization results is achieved. When the radiologists segmented the scar tissue and myocardium in both images, they computed the scar size and scar tissue ratio (%) for infarction patient. A strong correlation is found between the two scar sizes: the Pearson correlation coefficient (PCC) is 0.88, and the corresponding bias is -0.21 (limits of agreement) [2.2, -2.6] (P value= 0.31), as shown in Fig. 5.

#### TABLE II

<table>
<thead>
<tr>
<th></th>
<th>SpGAN</th>
<th>Spatiotemporal two-stream framework removed</th>
<th>Spatiotemporal feature pyramid removed</th>
<th>Spatiotemporal attention block block removed</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSIM</td>
<td>0.69±0.16</td>
<td>0.59±0.21</td>
<td>0.65±0.19</td>
<td>0.66±0.16</td>
</tr>
<tr>
<td>PSNR (dB)</td>
<td>20.84±1.31</td>
<td>20.72±1.53</td>
<td>20.21±1.40</td>
<td>19.97±1.33</td>
</tr>
<tr>
<td>NRMSE</td>
<td>0.31±0.06</td>
<td>0.37±0.11</td>
<td>0.35±0.09</td>
<td>0.34±0.08</td>
</tr>
<tr>
<td>PCC of Scar Size</td>
<td>0.88</td>
<td>0.82</td>
<td>0.86</td>
<td>0.79</td>
</tr>
</tbody>
</table>
TABLE III
PRODUCES THE HIGHEST PERFORMANCE WHEN COMPARED TO THAT OF VARIOUS BASELINE GAN FRAMEWORKS.

<table>
<thead>
<tr>
<th></th>
<th>SpGAN</th>
<th>MuTGAN</th>
<th>Pix2Pix</th>
<th>U-Net</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSIM</td>
<td>0.60±0.16</td>
<td>0.50±0.23</td>
<td>0.60±0.18</td>
<td>0.41±0.20</td>
</tr>
<tr>
<td>PSNR (dB)</td>
<td>20.84±1.31</td>
<td>17.34±1.56</td>
<td>20.03±1.34</td>
<td>17.10±1.41</td>
</tr>
<tr>
<td>NRMSE</td>
<td>0.31±0.06</td>
<td>0.47±0.08</td>
<td>0.38±0.05</td>
<td>0.52±0.10</td>
</tr>
</tbody>
</table>


D. Advantage of spatiotemporal feature pyramid and spatiotemporal attention block

Table III indicates that both the spatiotemporal feature pyramid and the spatiotemporal attention block effectively improve the quality of the converted infarction images and boost the infarction learning. The SpGAN improves the SSIM by 0.04, NRMSE by 0.04, and PCC of scar size by 0.02 compared with the version that removes the spatiotemporal feature pyramid. This improvement proves that SpGAN encodes rich semantics of cine MR images at all levels by combining coarse, semantically strong features with fine, semantically weak features. Moreover, the SpGAN improves the SSIM by 0.03, NRMSE by 0.03 and PCC of scar size by 0.09 compared with the version that removes the spatiotemporal attention block. This improvement proves that SpGAN successfully reinforces the details of the scar tissue by weighting these tissues during training.

E. The outperformance of our method than the existing CA-enhanced image converting methods

Table IV indicates the comparison between our method and few existing CA-enhanced image converting methods that mentioned in introduction (17, 15, 16). The comparison results shows that the SpGAN improves the SSIM by 0.19-0.51 and PCC of scar sizes by 0.31-0.67 than these three methods. All these improvements demonstrate that our method outperforms these three methods in infarction learning. Note that all CA-enhanced images generated by [17] is random and can not corresponds to patient’s non-enhanced images.

V. DISCUSSION

This paper reports a cross-modality data processing -based approach, named SpGAN that have great potential to apply the

TABLE IV
THE OUTPERFORMANCE OF OUR METHOD THAN THE EXISTING CA-ENHANCED IMAGE CONVERTING METHODS.

<table>
<thead>
<tr>
<th></th>
<th>SpGAN</th>
<th>Campello et al.</th>
<th>Hone et al.</th>
<th>Lau et al.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSIM</td>
<td>0.69</td>
<td>0.50</td>
<td>0.42</td>
<td>0.18</td>
</tr>
<tr>
<td>PCC of scar sizes</td>
<td>0.88</td>
<td>0.57</td>
<td>0.55</td>
<td>0.21</td>
</tr>
</tbody>
</table>
medical IoT in infarction learning, for the first time. Benefit from an exponential rise in IoT data, data processing has been proclaimed to be a key strategic technological trend in medical IoT [38], [39]. It integrates various efficient data analytical methods (especially artificial intelligence) to convert day-to-day medical data to smart data [40], [41], [42], [43]. It attains a better interpretation of data and formulates effective medical decisions to offer pervasive and personalized healthcare and medical services to people. Our SpGAN as a data processing approach designed for infarction learning leverages a well-trained deep learning framework to convert raw non-enhanced data from IoT-based MRI devices to contrast-enhanced data that is quite important to infarction diagnosis. It can embed a range of IoT-based MRI devices by building a unified edge-cloud computing platform [44], [45]. Thus, it allows a highly-efficient clinical way to remotely perform infarction observation and analysis to local patients, breaking through the limitations of time and space for timely infarction diagnosis and improving a patients diagnosis experience and treatment quality [46], [47]. It facilitates the infarction diagnosis repeatability and reduces the diversity between experienced radiologists and radiologists with limited experience, thereby improving the utilization rate of medical resources and the accuracy of diagnosis. Importantly, the application of our SpGAN and its IoT deployment offer a low-risk clinic manner to eliminate the potential high fatal risk in the current infarction learning workflow. Because of the limitations of contrast agent administration (e.g., using gadolinium chelate), the current ‘gold standard’ of infarction learning (i.e., contrast-enhanced data) may be fatal to patients with chronic kidney diseases. According to the US Renal Data System [38], [14], >40% of patients with chronic kidney disease also suffer from cardiovascular disease, and 20% of acute infarction patients are accompanied with chronic kidney disease. In contrast, our application requires only raw non-enhanced data from IoT-based devices, which is a part of the routine cardiac examination for function assessment. This makes our SpGAN produce a high clinical impact on the current infarction learning workflow, and thus promote the in-depth application and deployment of IoT in the medical field.

The future direction of our work include two parts. First, we plan to further introduce more medical IoT instruments with its image modality data to improve the quality of the converted images. Recent publications have proved that combining more image modalities (such as T2WI images or CT images) can prove more diagnostic information to achieve better performance. Second, we aim to integrate transfer learning to deploy medical IoT to more cardiac diseases. In this paper, we only study infarction learning. But in fact, basically all myocardial diseases can be learned through the spatiotemporal representation of the heart. In the next step, with the transfer learning, our method with its knowledge of the infarction can be designed and deployed for new medical IoT to more cardiac diseases such as myocarditis.

VI. CONCLUSION

In conclusion, a novel spatiotemporal two-streams generative adversarial network as a cross-modality data approach is proposed for deploying medical IoT in infarction learning. It can convert diagnostic valuable contrast-enhanced data directly from raw non-enhanced data of IoT-based MR devices and feed them to clinical routine. In real IoT-based remote infarction diagnosis, the radiologists compared the converted infarction images with the contrast-enhanced images of 230 subjects and found a strong correlation between them. A correlation of 0.90 was found in the 16-segment scar localization, and 0.84 was found in the scar ratio (%LV). These results demonstrate that SpGAN is an accurate clinical IoT tool to remotely assist patients, thereby improving the utilization rate of medical resources and the accuracy of diagnosis. It also facilitates patients who are limited in the use of CA by eliminating the use of CA and CA associated health risks, and reduces procedure cost.

REFERENCES

I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, 


A. Pumarola, A. Agudo, A. M. Martinez, A. Sanfeliu, and F. Moreno-

T. Van Laarhoven, “L2 regularization versus batch and weight normal-

S. Vesal, M. Gu, A. Maier, and N. Ravikumar, “Spatio-temporal multi-


Z. Gao, H. Xiong, X. Liu, H. Zhang, D. Ghista, W. Wu, and S. Li, “Robust estimation of carotid artery wall motion using the elasticity-


