Medical Image Analysis 71 (2021) 102037

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

# Super-Resolution of Cardiac MR Cine Imaging using Conditional GANs and Unsupervised Transfer Learning

Yan Xia[a,b,*], Nishant Ravikumar[a,b], John P. Greenwood[b], Stefan Neubauer[c],
Steffen E. Petersen[d,e], Alejandro F. Frangi[a,b,f]

[a] *Centre for Computational Imaging and Simulation Technologies in Biomedicine (CISTIB), School of Computing, University of Leeds, Leeds, UK*
[b] *Leeds Institute for Cardiovascular and Metabolic Medicine (LICAMM), School of Medicine, University of Leeds, Leeds, UK*
[c] *Oxford Center for Clinical Magnetic Resonance Research, Division of Cardiovascular Medicine, John Radcliffe Hospital, University of Oxford, Oxford, UK*
[d] *William Harvey Research Institute, NIHR Barts Biomedical Research Centre, Queen Mary University of London, London, UK*
[e] *Barts Heart Centre, St Bartholomew's Hospital, Barts Health NHS Trust, London, UK*
[f] *Medical Imaging Research Center (MIRC), Cardiovascular Science and Electronic Engineering Departments, KU Leuven, Leuven, Belgium*

ARTICLE INFO

*Article history:*
Received 8 November 2020
Revised 2 March 2021
Accepted 5 March 2021
Available online 6 April 2021

*Keywords:*
Cardiac MRI
Deep learning
Super-resolution
Conditional generative adversarial net
Optical flow
Conditional batch normalisation

ABSTRACT

High-resolution (HR), isotropic cardiac Magnetic Resonance (MR) cine imaging is challenging since it requires long acquisition and patient breath-hold times. Instead, 2D balanced steady-state free precession (SSFP) sequence is widely used in clinical routine. However, it produces highly-anisotropic image stacks, with large through-plane spacing that can hinder subsequent image analysis. To resolve this, we propose a novel, robust adversarial learning super-resolution (SR) algorithm based on conditional generative adversarial nets (GANs), that incorporates a state-of-the-art optical flow component to generate an auxiliary image to guide image synthesis. The approach is designed for real-world clinical scenarios and requires neither multiple low-resolution (LR) scans with multiple views, nor the corresponding HR scans, and is trained in an end-to-end unsupervised transfer learning fashion. The designed framework effectively incorporates visual properties and relevant structures of input images and can synthesise 3D isotropic, anatomically plausible cardiac MR images, consistent with the acquired slices. Experimental results show that the proposed SR method outperforms several state-of-the-art methods both qualitatively and quantitatively. We show that subsequent image analyses including ventricle segmentation, cardiac quantification, and non-rigid registration can benefit from the super-resolved, isotropic cardiac MR images, to produce more accurate quantitative results, without increasing the acquisition time. The average Dice similarity coefficient (DSC) for the left ventricular (LV) cavity and myocardium are 0.95 and 0.81, respectively, between real and synthesised slice segmentation. For non-rigid registration and motion tracking through the cardiac cycle, the proposed method improves the average DSC from 0.75 to 0.86, compared to the original resolution images.

## 1. Introduction

Cardiac Magnetic Resonance (MR) cine imaging allows structural and functional analysis of the heart, through accurate estimation of clinical parameters such as left ventricular (LV) volume and ejection fraction (EF), which are important predictors of clinical outcomes (Knauth et al., 2008). Due to its excellent reproducibility of quantitative measurements compared with other modalities, cardiac cine MR imaging is attractive for population imaging studies (Marwick et al., 2013) such as in the UK Biobank (Petersen et al., 2013), the German National Cohort (Bamberg et al., 2015), and the Canadian Alliance for Healthy Hearts and Minds (CAHHM) (Anand et al., 2016).

Although high-resolution (HR), isotropic cardiac MR images can facilitate better visualisation and more accurate assessment for complex cardiac morphology, 3D cine imaging of the heart is challenging due to relatively long acquisition and patient breath-hold times. Instead, the 2D b-SSFP (balanced steady-state free precession) sequence is regularly used in clinical practice as it provides excellent signal-to-noise ratio, good contrast between tissues and blood/vessels, and considerably reduces motion-induced signal dropout (Scheffler and Lehnhardt, 2003). However, typical 2D b-SSFP acquisitions consist of a few 2D short-axis (SAX) slices (gener-

---

* Corresponding author.
*E-mail address:* y.xia@leeds.ac.uk (Y. Xia).

https://doi.org/10.1016/j.media.2021.102037

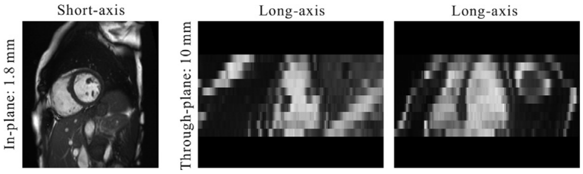

**Fig. 1.** An example of typical cardiac MR images from the UK Biobank with in-plane spatial resolution 1.8 × 1.8 mm, through-plane resolution 10 mm. From left to right, the three images display a short-axis (SAX) slice, two long-axis (LAX) slices, respectively. While in-plane spatial resolution is sufficient for image analysis, slice spacing is large due to acquisition constraints and hardly captures the detailed anatomy.

ally ranging between 8 and 12), resulting in highly anisotropic volumes with large through-plane spacing (e.g., 1.8 × 1.8 × 10 mm), as illustrated in Fig. 1. Such clinically acquired scans hamper not only visual interpretation, but also pose significant challenges to subsequent quantitative analyses. For instance, low through-plane resolution images make accurate non-rigid registration challenging to achieve, which is necessary for automatic cardiac strain analysis. Also, anisotropic 2D cine images may yield high variations in structural and functional assessment due to the difficulty of segmenting the myocardium near the apex and base of the ventricles. Therefore, these limitations have led to an interest in developing robust and efficient methods to improve the resolution of cardiac MR scans.

Traditional interpolation methods can be adapted to this scenario by using intensity/object-based interpolation. Intensity-based interpolation methods are essentially weighted averaging schemes of the input images, and thus yield blurring effects and unrealistic results. Object-based methods typically rely on image registration, where corresponding points between consecutive slices are found, and intermediate slices are subsequently interpolated. Such methods include the modified version of the control grid interpolation (CGI) (Frakes et al., 2008), multi-resolution registration-based slice interpolation (Leng et al., 2013), and higher-order spline-based interpolation (Horváth et al., 2017). However, registration-based slice interpolation methods have several limitations: First, the consecutive slices must have similar anatomical features. Second, the registration method must be able to estimate the correct transformation to match these similar features. Violation of either aspect yields false correspondence maps, which leads to incorrect interpolation results (cf. Fig. 2).

Conventional super-resolution (SR) techniques have also been applied to improve the resolution of cardiac MR images. These methods generally fall into two main categories: reconstruction through the combination of multiple acquired low-resolution (LR) orthogonal scans to achieve the higher resolution (ur Rahman and Wesarg, 2010; Gholipour et al., 2010), and example-based SR, which aims to up-sample LR images to their most likely HR version via knowledge of the relationship between HR and LR image features from example data (Manjón et al., 2010a; Rousseau et al., 2011; Konukoglu et al., 2013; Jog et al., 2014). The former not only requires additional acquisitions from multiple orthogonal views, but also depends on the quality of alignment of the images for fusion. Particularly, the non-rigid motion in cardiac imaging and the LR of the acquired images make accurate registration difficult to achieve, limiting the final effectiveness of such methods. The latter group of methods requires correspondences between LR and HR image patches from an example database, which may not always be available or feasible to acquire in clinical applications.

Similarly, one can adopt deep convolutional neural networks (CNN) to learn end-to-end mapping between the LR/HR images directly (Dong et al., 2014; Oktay et al., 2016; Basty and Grau, 2018; Chaudhari et al., 2018). There were also SR methods based on generative adversarial nets (GANs) that have demonstrated to boost the performance for the super-resolved MRI or CT images. For instance, (Chen et al., 2018b) combined a light-weighted densely connected network and a GAN to provide the state-of-the-art performance while keeping the model smaller and faster. (Lyu et al., 2019) proposed a GAN-CIRCLE network that was constrained by the identical, residual, cycle learning ensemble and achieved two-fold resolution enhancement for MRI and CT. A multi-scale GAN with lesion focused SR was also studied to achieve stable and efficient training and improve perceptual quality of super-resolved results (Zhu et al., 2019). (Wang et al., 2020) suggested an enhancement to tackle GAN-based 3D SR by introducing a residual-in-residual dense block (RRDG) generator that is both memory efficient and achieves state-of-the-art performance. (Kudo et al., 2019) proposed to condition the discriminator on the body parts and realised CT image super resolution for different body parts with a single 3D GAN. However, acquiring either the appropriate HR, near isotropic 3D cardiac MR images or multiple 2D LR stacks in different orientations is often infeasible and impractical. Thus, here our discussion only focused on the typical clinical scenario where only sparsely-sampled SAX images were available.

There also exist methods that attempt to fill in the missing slices by exploring redundant, relevant details and capturing repetitive anatomical structure in a scan or in clinical image collections (Manjón et al., 2010b; Plenge et al., 2013; Dalca et al., 2018). Recently, (Zhao et al., 2020) used in-plane HR information to restore LR through-plane slices. They first trained a regression model to generate in-plane image patches at the original resolution acquired, from LR image patches generated by filtering the original image patches. Subsequently, they applied the trained regression model in the through-plane direction to generate self-supervised SR images. However, these methods were designed and validated on images involving almost no organ motion, such as brain MRI and hence may not deal with motion often presented in dynamic cardiac cine MR acquisitions, such as inter-slice motion and misalignment.

Alternatively, video frame interpolation, which has been extensively studied in the computer vision (CV) domain, may be applied in this scenario. The most common approach to video interpolation is based on optical flow (Herbst et al., 2009; Liu et al., 2017; Jiang et al., 2018; Niklaus and Liu, 2018; Bao et al., 2019b). These methods first estimate object motion through optical flow using CNN models and then perform frame warping to synthesise pixels using reference frames. Instead of estimating a temporal motion field in the adjacent frame sequences, the same underlying concept can also be adapted to derive spatial motion/transformation from two consecutive slices in the through-plane direction, and synthesise the corresponding intermediate slice. While achieving impressive performance, these methods still have several limitations: First, the frames synthesised from optical flow suffer from blurriness, as inaccurate flow estimation from frames far apart and bilinear warping blend neighbouring pixels, due to the sub-pixel shifts. Second, the optical flow estimation can struggle to maintain accuracy in the presence of large motion patterns, leading to ghost or overlaid artefacts. These artefacts are even more pronounced in cardiac MR images as the anatomical structures change substantially between consecutive slices (cf. Fig. 2). Recently, (Bao et al., 2019a) suggested considering the contribution of each optical flow based on the depth information and compensated for the issue that multiple flow vectors may overlap at the same position. Also, to reduce artefacts around motion boundaries, the authors used an adaptive warping layer to warp the input frames, contextual features, and depth maps based on the estimated flows and local interpolation kernels, and refined the output frame through a CNN-based, cascaded model. However, although this depth-aware frame interpolation can generate frames with precise object shapes, the synthesis network tends to generate blurry images with fewer details, since

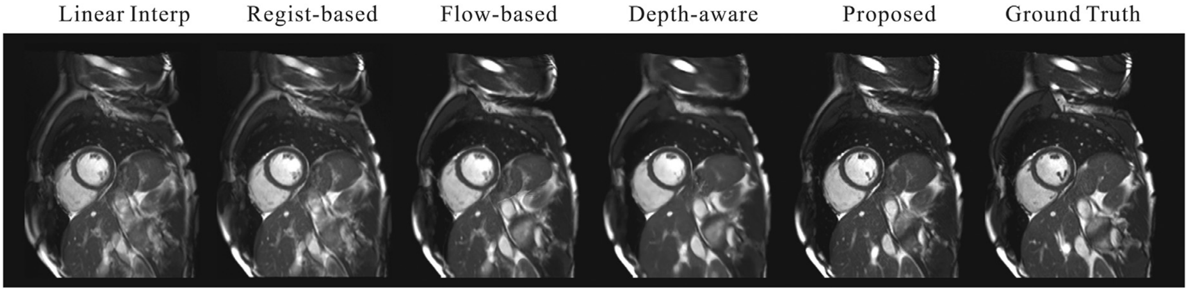

**Fig. 2.** A comparison of the reference slice and the synthesised slices between the linear interpolation, registration-based interpolation (Horváth et al., 2017), optical flow-based frame interpolation (Jiang et al., 2018), depth-aware flow interpolation (Bao et al., 2019a) and the proposed SR approach for the UK Biobank cardiac MR dataset.

the generic CNN cannot capture the multi-modal distribution of real images, as can be seen in Fig. 2.

Our goal was to reconstruct high-resolution, 3D isotropic cardiac MR images, from single SAX stacks of sparsely sampled 2D slices. Several aspects were challenging: First, the through-plane spacing of the SAX image is approximately 8 times the in-plane spatial resolution (e.g. 10 mm vs 1.2 mm), whilst the up-sampling factor of classic SR algorithms is usually around 2-4. Second, we focused on data typically acquired in a clinical setting, i.e. neither the relationships between HR/LR image features, nor additional scans from other orthogonal views, which most SR algorithms rely on, were available. No ground-truth was available for training, as those intermediate slices we aim to synthesise, were not acquired in the standard clinical 2D b-SSFP imaging protocol due to acquisition constraints. Finally, cardiac images exhibit a significant degree of variability in terms of orientation, anatomy and motion. For instance, inter-slice misalignment may exist within a SAX stack, due to variations in the position of the diaphragm during breath-hold acquisitions.

In this work, we proposed a novel, robust method based on recent techniques of conditional GANs (Goodfellow et al., 2014; Mirza and Osindero, 2014; Isola et al., 2017). We showed that with dedicated loss functions, adversarial optimisation, and an auxiliary image generated by a state-of-the-art optical flow method (Bao et al., 2019a), our network can be trained end-to-end using unsupervised transfer learning. Here, transfer learning indicates that the state-of-the-art optical flow model used to generate guided images was pretrained on the “labelled” video clips and video frames in the CV domain. The proposed framework effectively incorporates visual properties and relevant textures of input images and can synthesise HR anatomically plausible 3D cardiac MR images, consistent with the available slices. The main contributions of our approach are:

(1) A novel deep conditional GAN architecture was proposed to enable HR, 3D isotropic cardiac MR reconstructions, using single sparsely-sampled image stacks. The method does not require the corresponding HR scans or multiple LR scans.

(2) To synthesise visually appealing cardiac MR images and facilitate accurate quantitative measurements, we adopted a dedicated generator and discriminator in this work. The generator contains residual blocks, where all normalisation layers are conditioned and modulated with two consecutive input images to ensure that visual properties and relevant texture details are effectively propagated through the network. A multi-scale discriminator was employed to ensure recovery of both global and local spatial features.

(3) To maintain precise object shapes and motion consistency, we adopted a state-of-the-art optical flow interpolation technique in the CV context to produce an intermediate auxiliary image. This was used to guide image generation by matching its features with the synthesised image. We proposed a composite loss function to train the network in an unsupervised transfer learning fashion.

(4) We comprehensively analysed the model by evaluating its performance across a large cohort of subjects at, a single cardiac phase and across multiple time points in the cardiac cycle. We demonstrated that subsequent image analyses including ventricle segmentation, cardiac quantification and non-rigid registration can benefit from the proposed method by generating super-resolved, isotropic cardiac MR images.

The paper is organised as follows. Section II introduces the generative model and learning algorithm. Section III describes the experiments conducted to validate the model, and Section IV presents the results of qualitative and quantitative analyses conducted to evaluate model performance. Then, we discuss essential characteristics of our model in contrast with the state-of-the-art, and its relevance in real clinical scenarios in Section V, before providing concluding remarks for our work in Section VI.

## 2. Methods

In this section, we present the generative model for intermediate slice synthesis, the adversarial learning algorithm, and describe the image synthesis pipeline. Overview of the method: First, a generative network takes the two consecutive, down-sampled images $x_{z\pm1}$ as conditioning input to synthesise an intermediate cardiac MR slice. Then, a multi-scale discriminator distinguishes the generated samples from the real images at different scales, and simultaneously matches the features of the inferred slice with those of the real slices, $x_{z\pm1}$ and an auxiliary image $\hat{x}_z$ generated from the optical flow-based interpolation, using a hybrid feature matching loss. During inference, all the original slices are maintained and we fill in the in-between slices with the newly synthesised ones (using every two consecutive slices) to form the new image volume. Then, we repeat this up-sampling process on the new image volume to achieve higher resolution in the through-plane direction. The network architecture is illustrated in Fig. 3.

### 2.1. Generative Adversarial Networks (GANs)

A GAN consists of two networks: the generator $G$ and the discriminator $D$ (Goodfellow et al., 2014). In an image synthesis, $G$ builds a mapping function that maps a random noise vector $z$ to an output image $x$ from a target distribution $p(x)$, and $D$ outputs a single scalar that represents the probability that a sample $s$ is drawn from $p(x)$. On the one hand, $G$ learns to synthesise realistic-looking images that cannot be distinguished from real images by an adversarially trained $D$. On the other hand, $D$ tries to distinguish real images from the generated ones. The minimax objective

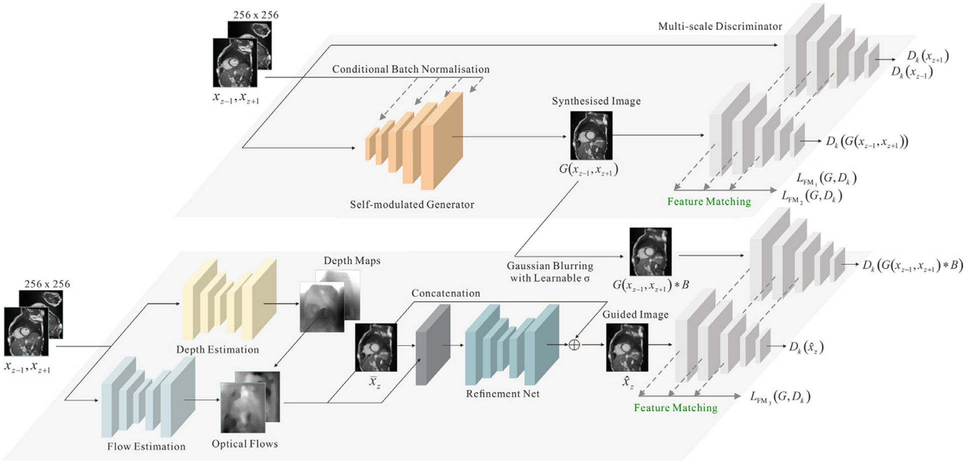

**Fig. 3.** Structure of the proposed SR network for generating high-resolution, anatomically realistic images consistent with the available slices. The generative network takes the two consecutive, down-sampled images $x_{z\pm1}$ as conditioning input to synthesise a cardiac MR slice $G(x_{z-1}, x_{z+1})$ at the middle position. The multi-scale discriminator distinguishes the generated samples from the real images at different scales, and simultaneously matches the features of the inferred slice with those of the real slices $x_{z\pm1}$ and an auxiliary image $\hat{x}_z$ generated from the optical flow-based interpolation using different feature matching loss terms.

for GANs can be formulated as follows[1]:

$$\min_G \max_D \mathcal{L}_{GAN} = \min_G \max_D \mathbb{E}_x[\log D(x)] + \mathbb{E}_z[\log(1 - D(G(z)))]. \tag{1}$$

To improve network stability during training, the negative log-likelihood in $\mathcal{L}_{GAN}$ can be replaced by a squared loss function (Mao et al., 2017):

$$\mathcal{L}_{LSGAN} = -\mathbb{E}_x\left[(D(x) - 1)^2\right] - \mathbb{E}_z\left[D(G(z))^2\right]. \tag{2}$$

### 2.2. Image-Conditional Generative Adversarial Networks

Instead of learning a mapping from random noise $z$, (Isola et al., 2017) proposed an image-to-image translation model (also known as the pix2pix model) that learns a mapping from statistically-dependent observed source images $x$ to target images $y$. Such a method can be regarded as image-conditional GAN. The optimisation of the $G$ and $D$ can be reformulated as:

$$\mathcal{L}_{cGAN} = -\mathbb{E}_{xy}\left[(D(x, y) - 1)^2\right] - \mathbb{E}_x\left[D(G(x))^2\right]. \tag{3}$$

It has been shown an effective strategy to integrate the traditional pixel-wise loss (e.g., L1 or L2 distance between the ground truth and generated images) into the GAN objective function for boosting image generation tasks:

$$\mathcal{L}_{L1} = \mathbb{E}_{xy}[\|y - G(x)\|_1]. \tag{4}$$

Then, the final loss function combines Eq. (3) and (4). However, the original pix2pix models may be unstable and prone to failure for synthesising images that contain fine structural details and rich quantitative information, which is essential for medical images. Our generative model is inspired by recent GAN architectures to refine the previous models and aims to generate anatomically plausible cardiac MR images which preserves fine structural details. The architecture of the generator and discriminator will be discussed in subsequent sections.

### 2.3. Generator with Self-modulated Normalisation

Our generator uses a pre-activation ResNet architecture (He et al., 2016) that was implemented in recent popular GAN models (Karras et al., 2018; Zhu et al., 2017; Isola et al., 2017; Miyato et al., 2018; Zhang et al., 2019). As shown in Fig. 4, the generator comprises of several residual blocks, and nearest neighbour up-sampling layers. Each residual block contains two convolutional layers with the skip connection, and a learned residue of input is added to the output to ensure the characteristics of original features are retained.

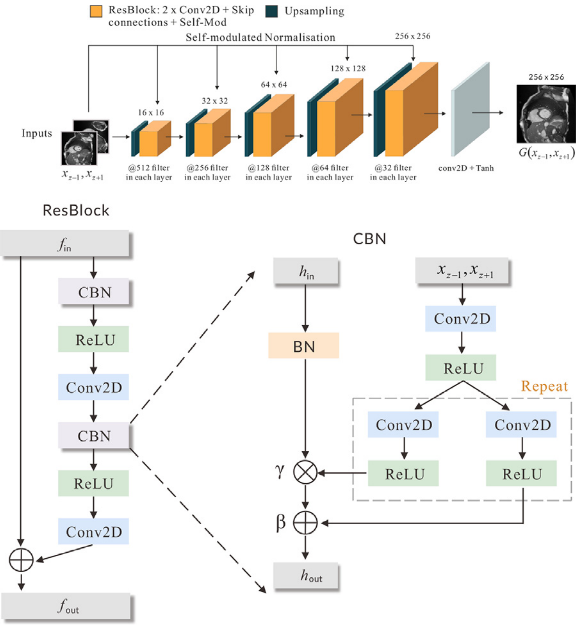

**Fig. 4.** Structure of the proposed generator. The generator follows a full pre-activation ResNet architecture that consists of residual blocks, followed by nearest neighbour up-sampling layers. Conditional batch normalisation (CBN) is implemented in each ResBlock so feature maps are first normalised to zero mean and unit deviation, followed by modulation/de-normalisation using a learned transformation whose parameters are inferred from two input cardiac MR slices, facilitating that fine details are propagated throughout the network.

Additionally, we were inspired by the concept of conditional batch normalisation (CBN) that has been recently adopted in previous studies (De Vries et al., 2017; Miyato and Koyama, 2018; Zhang et al., 2019; Chen et al., 2018a; Park et al., 2019). CBN sug-

---

[1] Here, we denote $\mathbb{E}_{x\sim p_{\text{data}}(x)}$ as $\mathbb{E}_x$ and $\mathbb{E}_{z\sim p_z(z)}$ as $\mathbb{E}_z$ for simplicity.

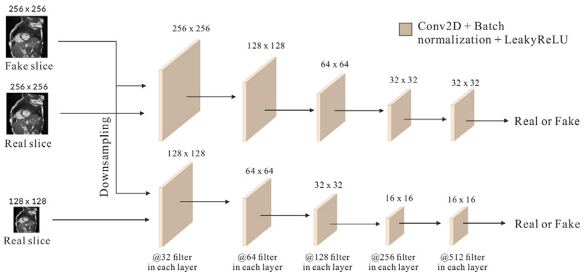

**Fig. 5.** The multi-scale discriminative network that operates at different image scales (256 × 256 and 128 × 128) and helps to capture both global and local spatial features.

gests an effective guidance strategy to incorporate additional conditioning information (such as labels, embedding, or masks) into image synthesis through batch normalisation and enables an image to be translated from one domain into another while consistently respecting the constraints specified by conditioning information. Specifically, in CBN layers, the extracted features from the previous layers are first normalised to zero mean and unit standard deviation. Then, the normalised features are modulated using a learning-based affine transformation with scale and shift parameters inferred from the external conditioning data. Significant improvements in synthesised image quality have been demonstrated for a wide variety of settings (Ulyanov et al., 2017; Huang and Belongie, 2017; Chen et al., 2018a; Park et al., 2019).

Our CBN scheme is close to the self-modulated normalisation (Chen et al., 2018a) and the spatially-adaptive normalisation (Park et al., 2019) that was used for style transfer. However, our method differs from the previous methods in application-specific design: we aim to produce image reconstruction that effectively incorporates visual properties and quantitative information from the adjacent input slices by consistently encoding them into the image generation pathway through the normalisation layers, to capture spatial features from all relevant regions near the synthesised position. By doing so, we can generate a slice that closely resembles the appearance of the input reference slices. Specifically, in the batch normalisation setting, input feature batch $h_{b,w,h,c} \in \mathbb{R}^{B\times W\times H\times C}$ ($b \in B, w \in W, h \in H$, and $c \in C$ denote the batch size, width, height, and channel of the feature map, respectively) is normalised in a channel-wise manner:

$$\begin{aligned} h'_{b,w,h,c,} = {} & \gamma_{w,h,c}(x_{z-1}, x_{z+1}) \times \frac{h_{b,w,h,c} - \mu_c}{\sigma_c + \epsilon} \\ & + \beta_{w,h,c}(x_{z-1}, x_{z+1}), \end{aligned} \tag{5}$$

with

$$\mu_c = \frac{1}{N}\sum_{b,w,h} h_{b,w,h,c}, \quad \sigma_c^2 = \frac{1}{N}\sum_{b,w,h}\left(h_{b,w,h,c} - \mu_c\right)^2, \tag{6}$$

where $N = B \times W \times H$, $\epsilon$ is a small number to avoid division by zero, $x_{z\pm1}$ denotes two consecutive input slices, $\gamma(\cdot)$ and $\beta(\cdot)$ are spatial dimension-dependent functions that can be formulated as a simple CNN model. The modulation parameters of all CBN layers within the generator were learned simultaneously through the GAN training.

### 2.4. Multi-scale Discriminator

In this work, we employed a multi-scale discriminator (Durugkar et al., 2017; Nguyen et al., 2017; Wang et al., 2018) that operates at two scale levels, as shown in Fig. 5. All the input images to the first discriminator are down-sampled by a factor of 2 for the second discriminator, and the network is trained in a multi-task fashion. The rationale for applying this two-scale image pyramid structure is: the discriminator with the larger receptive field yields a global view of the image and can guide the generator to synthesise globally coherent images, whilst the discriminator with a smaller scale can encourage the generator to capture local finer details. Batch normalisation and LeakyReLU with slope 0.2 are applied for all the layers in the discriminator.

### 2.5. Depth-Aware Optical Flow Interpolation

A conventional conditional GAN can synthesise an image that resembles the appearance of the two input images, but it is hard to control the movement patterns to ensure that the synthesised image lies precisely between and central to the neighbouring slices. To overcome this issue, we were inspired by the recent advances in video frame interpolation in (Jiang et al., 2018; Bao et al., 2019a; 2019b) and adapted a depth-aware flow-based interpolation approach (Bao et al., 2019a) to generate an auxiliary image to guide adversarial learning.

Specifically, we first computed the bi-directional optical flows between two consecutive input images using a state-of-the-art optical flow model. Let $F_{z+1\to z-1}$ and $F_{z-1\to z+1}$ denote the bi-directional optical flow from $x_{z+1}$ to $x_{z-1}$ and from $x_{z-1}$ to $x_{z+1}$, respectively. Then, the input images can be warped by the estimated intermediate optical flows to synthesise the intermediate image $\breve{x}_z$ as:

$$\breve{x}_z = \frac{1}{2}g(x_{z-1}, F_{z\to z-1}) + \frac{1}{2}g(x_{z+1}, F_{z\to z+1}), \tag{7}$$

with the approximated intermediate optical flows:

$$F_{z\to z-1} \cong -\frac{1}{2}F_{z-1\to z+1} \quad F_{z\to z+1} \cong -\frac{1}{2}F_{z+1\to z-1}, \tag{8}$$

where $g(\cdot,\cdot)$ is a backward warping function, which can be implemented using bilinear interpolation (Zhou et al., 2016). As there may exist multiple flow vectors projected to the same location in the intermediate frame, a depth-aware flow interpolation was proposed in (Bao et al., 2019a) and Eq. (8) can be reformulated as:

$$\bar{F}_{z\to z-1} \cong -\frac{W_{z-1}\cdot F_{z-1\to z+1}}{2\|W_{z-1}\|} \quad \bar{F}_{z\to z+1} \cong -\frac{W_{z+1}\cdot F_{z+1\to z-1}}{2\|W_{z+1}\|}, \tag{9}$$

where the weights $W_{z-1}$ and $W_{z+1}$ are the reciprocal of the corresponding depth maps for $x_{z-1}$ and $x_{z+1}$, respectively. Eq. (9) implies the projected flows tend to sample the closer objects and reduce the contribution of the pixels yielding larger depth values. Inserting $\bar{F}_{z\to z-1}$ and $\bar{F}_{z\to z+1}$ into Eq. (7), we can obtain $\bar{x}_z$. To reduce artefacts around motion boundaries that cause poor image synthesis, the initial approximation $\bar{x}_z$ is refined through the cascaded CNN architecture as adopted in (Jiang et al., 2018; Bao et al., 2019a) to obtain the intermediate guidance image $\hat{x}_z$ for our model.

### 2.6. Optimisation

In this section, we present a new loss function to train the network adversarially, in an unsupervised transfer learning manner. By adopting the squared loss function and the multi-scale discriminator, our GAN training loss can be formulated as:

$$\begin{aligned} \mathcal{L}_{GAN}(G, D_k) = {} & \min_{G}\max_{D_1,D_2}\sum_{k=1,2} -\mathbb{E}_x\left[(D_k(x_{z-1}) - 1)^2\right] \\ & -\mathbb{E}_x\left[(D_k(x_{z+1}) - 1)^2\right] - \mathbb{E}_x\left[D_k(G(x_{z-1}, x_{z+1}))^2\right]. \end{aligned} \tag{10}$$

Instead of using pixel-wise loss in Eq. (4) and minimising the L1 distance to the ground truth, a feature matching loss was employed to optimise the GAN to match the statistics of feature representations between the two input slices and synthesised slice, in multiple intermediate layers of $D_k$:

$$\mathcal{L}_{FM_1}(G, D_k) = \mathbb{E}_x\sum_{i=1}^{T}\frac{1}{N_i}\left[\|D_k^i(x_{z-1}) - D_k^i(G(x_{z-1}, x_{z+1}))\|_1\right], \tag{11}$$

$$\mathcal{L}_{FM_2}(G, D_k) = \mathbb{E}_x \sum_{i=1}^{T} \frac{1}{N_i}\left[\|D_k^i(x_{z+1}) - D_k^i(G(x_{z-1}, x_{z+1}))\|_1\right], \qquad (12)$$

where $i$ means the $i$th layer features in $D$, $N_i$ is the number of features in each layer, $T$ is the total number of layers.

Then, we introduced another feature matching term that compares the synthesised image with a guidance image $\hat{x}_z$ estimated from the depth-aware optical flow interpolation in the sparse representation:

$$\mathcal{L}_{FM_3}(G, D_k) = \mathbb{E}_x \sum_{i=1}^{T} \frac{1}{N_i}\left[\|D_k^i(\hat{x}_z) - D_k^i(G(x_{z-1}, x_{z+1}) * B)\|_1\right], \qquad (13)$$

where $B$ represents a Gaussian smoothing operation with a learnable standard deviation $\sigma$.

The motivation of this hybrid feature matching loss is: The configuration of Eq. (11) and (12) aims to generate anatomically plausible images resembling the appearance of the two input images, whilst Eq. (13) encourages the generated intermediate image to yield motion consistency and smooth transition to the neighbouring slices. Our final joint objective combined both GAN loss and feature matching loss:

$$\mathcal{L}_{SR} = \min_{G}\Bigg(\Bigg(\max_{D_1, D_2} \sum_{k=1,2} \mathcal{L}_{GAN}(G, D_k)\Bigg) + \lambda_1 \sum_{k=1,2} \mathcal{L}_{FM_1}(G, D_k)$$
$$+ \lambda_2 \sum_{k=1,2} \mathcal{L}_{FM_2}(G, D_k) + \lambda_3 \sum_{k=1,2} \mathcal{L}_{FM_3}(G, D_k)\Bigg), \qquad (14)$$

where $\lambda_1$, $\lambda_2$ and $\lambda_3$ control the weighting of the feature matching loss to GAN loss.

## 3. Experimental Setup

### 3.1. Dataset

Cardiac MR images from the UK Biobank (UKBB) were used to train and validate the proposed method. Images were acquired using a clinical wide bore 1.5T MR system (MAGNETOM Aera, Syngo Platform VD13A, Siemens Healthcare, Erlangen, Germany) equipped with an 18 channel anterior body surface coil (45 mT/m and 200 T/m/s gradient system). 2D cine b-SSFP SAX image stacks were acquired with the following acquisition protocol: in-plane spatial resolution 1.8 × 1.8 mm, slice thickness 8 mm, slice gap 2 mm, image size 198 × 208. The number of slices in the SAX stack typically ranges between 8 and 12. Each slice was acquired at 50 cardiac phases. Further acquisition details can be found in (Petersen et al., 2015). In this work, we focused on the cardiac SAX b-SSFP cine MR datasets for which the ground-truth intermediate slices that correspond to the desired high-resolution 3D isotropic images, were not available.

### 3.2. Network Training

Our approach adopted a least-squares GAN (refer to Eq. (2)), and was trained on 3,000 pairs of two consecutive slices (each pair taken from a different subject at a random slice position) from sparsely-sampled UKBB cardiac MR images within a single cardiac phase (end-diastolic) for 50 epochs. The aim was to synthesise the unseen intermediate slice. The training uses the Adam optimiser with an initial learning rate of $2 \times 10^{-4}$, for both the generator and discriminator. The decay rates of the first and the second momentum of the gradient estimates were set to 0.5 and 0.999, respectively. All the 2D image slices were resized to 256 × 256. The method requires no HR scans or intermediate ground-truth images and can be trained in an unsupervised transfer learning fashion. The relative weighting factors of the feature matching loss to GAN loss in Eq. (14) were initialised as $\lambda_1 = 0.33$, $\lambda_2 = 0.33$ and $\lambda_3 = 0.33$ and adjusted to $\lambda_1 = 0.2$, $\lambda_2 = 0.2$ and $\lambda_3 = 0.6$ after 15 epochs. Instead of performing standard batch normalisation, in the generator we first normalised all intermediate feature maps to zero mean and unit standard deviation, followed by de-normalising/modulating features using a learned transformation whose parameters are inferred from the input cardiac MR slices. For intermediate slice synthesis, we utilised the two adjacent slices as input (see Eq. (5)) to ensure the normalisation is spatially dependent.

To synthesise an intermediate, auxiliary image to guide image generation and used by the third featuring match loss term in Eq. (13), we fine-tuned the depth-aware flow interpolation model that was pre-trained on 51,312 Vimeo90K datasets (Xue et al., 2019) using 3,000 UKBB cardiac MR image triplets at the original low slice resolution (10 mm) for 30 epochs. The guidance image is only utilised for the training stage. For the inference, we do not need the computation of the bi-directional optical flows and depth maps as the intermediate slice can be directly synthesised from two input images through the generative model.

### 3.3. Competing Methods

To demonstrate the advantages of the proposed method, we compared it to two conventional intensity-based and registration-based interpolation methods. The intensity-based method simply computes the weighted average of adjacent slices using bicubic interpolation in the slice direction to obtain isotropic voxels. The registration-based slice interpolation approach (Horváth et al., 2017) first uses a symmetric similarity measure to perform structure registration, to calculate displacement fields between neighbouring slices. Then, along every correspondence point trajectory, the displacement fields are utilised to calculate a high order intensity interpolating spline for structural motion. The algorithm was executed using the recommended parameter settings described in their work.

We also compared our method to a state-of-the-art optical flow interpolation approach (Jiang et al., 2018) and the depth-aware flow interpolation approach (Bao et al., 2019a). The optical flow-based model was pre-trained using two datasets comprising of 1,132 video clips and 376K individual video frames from a great variety of scenes. We fine-tuned both models with 3,000 UKBB cardiac MR image triplets at the current through-plane resolution (10 mm spacing) for 30 epochs, to predict the middle slice from pairs of the first and third slices.

### 3.4. Evaluation Design

We conducted several experiments to assess the accuracy and robustness of the proposed SR method. First, images generated using the proposed method, from UKBB data comprising subjects at both the end-diastolic (ED) and end-systolic (ES) cardiac phase, were qualitatively assessed. We analysed our method's performance at synthesising slices across different resolution levels, i.e. from the original resolution with the slice spacing of 10 mm to 16x up-sampled HR image stacks where the slice thickness eventually reduces to 0.625 mm. We also investigated the case where one slice is removed from original sparsely-sampled SAX images, before using the proposed method to synthesise the missing slices. Note that, only in this scenario, the ground-truth images were available for comparison. All the results obtained using our approach were compared with those of the other approaches investigated.

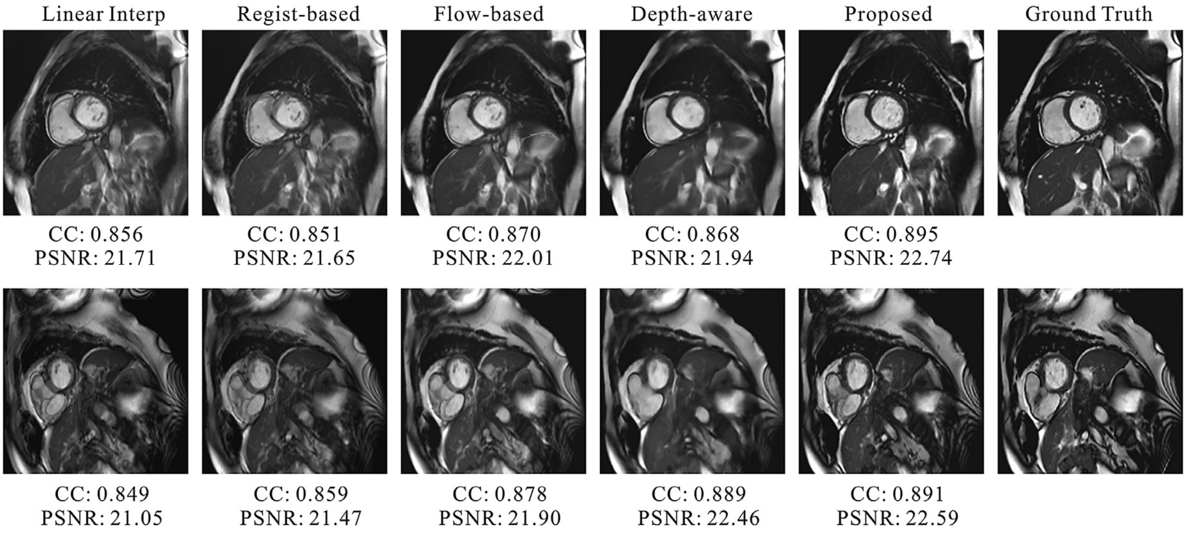

**Fig. 6.** Qualitative comparison of the reference slice and the synthesised slices between the linear interpolation, registration-based interpolation, optical flow based frame interpolation, depth-aware flow interpolation and the proposed SR approach for two cardiac MR datasets at the original slice resolution 10 mm. The comparison is augmented by the computed correlation coefficients (CC) and PSNR shown below each figure.

We investigated the impact of the proposed SR approach on typical cardiac MR image analyses, including biventricular segmentation, cardiac quantification and temporal non-rigid registration across the full cardiac cycle. In these study groups, we adopted the bicubic interpolation method as the baseline for comparison since it has been commonly used to produce near isotropic cardiac MR images in these applications. First, for automated segmentation, we adopted a state-of-the-art, 2D CNN-based method proposed in (Bai et al., 2018) that can take a SAX image and predict a pixel-wise image segmentation through fully convolutional layers. We retrained this CNN model on 4,000 cardiac MR datasets in total split into three subsets of 3100, 500 and 400 for training, validation and testing, respectively. Manual image annotations of LV endocardium, LV myocardium, and RV endocardium were taken from (Petersen et al., 2017). All the 2D slice images were cropped to a $192 \times 192$ matrix and intensity normalised to the range of [0,1]. To quantify segmentation accuracy, three common metrics were used, namely Dice similarity coefficient (DSC), mean contour distance (MCD) and Hausdorff distance (HD). The DSC computes the overlap between two segmentations $A$ and $B$ as:

$$DSC = \frac{2|A \cap B|}{|A| + |B|}. \tag{15}$$

DSC is in the range between 0 and 1, with a higher value indicating a better match in two segmentation results. The MCD and HD metrics evaluate the mean and maximum distance between segmentation contours $\partial A$ and $\partial B$ as:

$$MCD = \frac{1}{2|\partial A|} \sum_{p \in \partial A} d(p, \partial B) + \frac{1}{2|\partial B|} \sum_{q \in \partial B} d(q, \partial A), \tag{16}$$

$$HD = \max \left( \max_{p \in \partial A} d(p, \partial B), \max_{q \in \partial B} d(q, \partial A) \right), \tag{17}$$

where $d(p, \partial)$ denotes the minimal distance from point $p$ to contour $\partial$. Lower distance metric values indicate a better agreement.

We also evaluated the accuracy of clinical measures such as ventricular volume, mass and wall thickness derived from the automated segmentations. The LV and RV volumes (ml) indicate the volume of blood in the LV and RV, and are computed by summing up the number of voxels in the segmentation results and multiplying by the pixel spacing. The LV mass was calculated by multiplying the LV myocardial volume by the known density of 1.05 g/mL. Wall thickness is expressed in mm and measures the distance between the endocardial and epicardial walls and is often used as a biomarker for quantifying regional dysfunction.

To investigate the impact of the SR method on non-rigid registration and motion tracking, we used a diffeomorphic non-linear registration algorithm (Vercauteren et al., 2007) that finds an invertible one-to-one mapping between two images. Estimation was performed by co-registering the 3D heart region throughout the whole cardiac sequence with 50 time phases on 100 test sets (i.e., 5,000 image stacks). A recursive strategy was adopted to assess accumulated errors during registration: First, pair-wise, frame-to-frame spatial transformations $\varphi_{v_t \to v_{t-1}}$ ($v_t$ denotes the image stack at the time point $t$) were computed by aligning two consecutive frames in a forward, temporal direction, followed by estimating the frame-to-reference transformation $\varphi_{v_t \to v_0}$ through compositions $\varphi_{v_t \to v_{t-1}} \circ \varphi_{v_{t-1} \to v_{t-2}} \cdots \varphi_{v_1 \to v_0}$. Similarly, the backward pass was estimated in reverse order. Then, both deformations were combined using a weighting function, according to the distance between each frame and the reference frame. Finally, we used the combined frame-to-reference transformation to warp segmentation masks at different cardiac time points $t \in [1, 49]$ to the ED phase $t = 0$ and estimated the accuracy using the DSC, MCD and HD metrics.

## 4. Results

### 4.1. Image Quality Assessment

First, a qualitative comparison of the synthesised slices can be seen in Fig. 6 including, linear interpolation, registration-based interpolation, optical flow based interpolation, depth-aware flow interpolation and the proposed SR approach for two examples of the UKBB dataset at the original 10 mm slice thickness. Here, the corresponding ground-truth images were available. The comparison is also augmented by the computed correlation coefficients (CC) and peak signal-to-noise ratio (PSNR) shown below each figure. We can see that the simple linear interpolation yields uncertain and highly overlapping shapes in the interpolated slices. Results of the registration-based method are similar to that of the linear interpolation. This is because considerable dissimilarity between adjacent cardiac MR slices yields false correspondence maps during registration, and leads to incorrect interpolation results. Al-

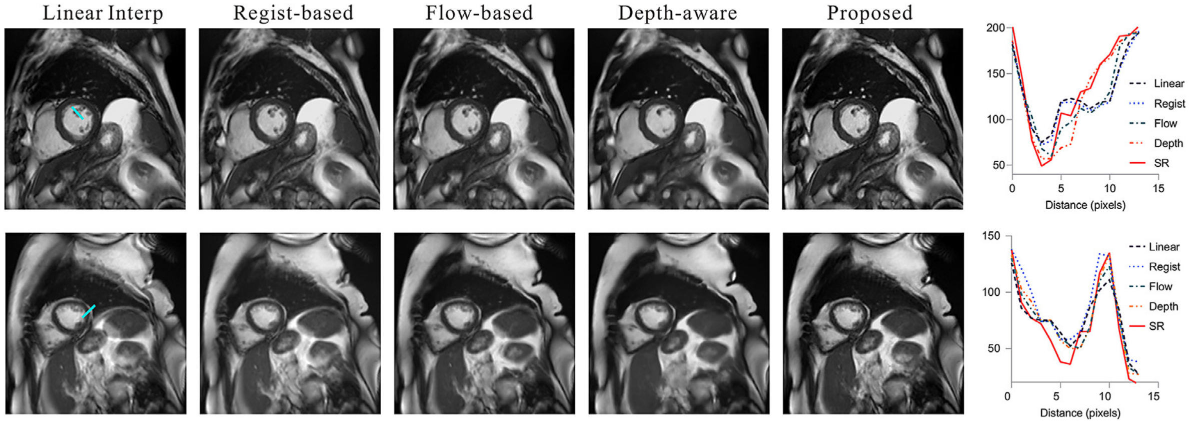

**Fig. 7.** Qualitative comparison of the synthesised slices between the linear interpolation, registration-based interpolation, optical flow-based frame interpolation, depth-aware flow interpolation and the proposed SR approach for two cardiac MR datasets at the slice resolution 5 mm. The ground-truth images are not available at this resolution level. Intensity line profiles drawn across the LV (indicated by the green line) are shown on the right column, demonstrating the resolution improvement in the results of the proposed SR method.

though the optical flow-based method outperforms the traditional interpolation method, it could not eliminate the ghost effect due to large motion variations. The depth-aware flow interpolation reduces artefacts around motion boundaries that cause inadequate image synthesis, but produces blurred images, which is a trade-off for the refinement around motion boundaries. In contrast, our approach preserves fine structural details and realistic textures and is most comparable in terms of visual quality to the ground-truth image. Besides qualitative improvements, the CC and PSNR values of the proposed method are also improvements over traditional and state-of-the-art interpolation methods.

Fig. 7 shows the synthesised intermediate slice from neighbouring slices with 5 mm spacing (versus 1.8 mm in-plane). Similar to the previous experiment, the proposed SR method exhibits better visual image quality, recovers more details and enhanced edges, such as fat, papillary muscles and small vessels. Note that the ground-truth images were not available at this resolution. Intensity line profiles drawn across the LV (indicated by the green line) demonstrate the resolution improvement in the results of the proposed method, with anatomical structures such as the myocardial wall and papillary muscles appearing sharper.

Important differences are also observed in the through-plane direction, i.e. in the long-axis (LAX) orientations, as depicted in Figs. 8, 9 and 10. Fig. 8 illustrates the progressive resolution improvement from the LAX view, where the native cardiac MR images were up-sampled with the scale factor of 2, 4, 8 and 16 by employing the proposed method in a recursive manner. The right column shows the zoomed-in, cropped ROI region in the LV at different up-sampling levels. We can see that the anatomical structures and edges are not visible in the original resolution due to severe staircasing effects but are smoothly reconstructed by the proposed SR approach (see noticeable differences near the basal, apical region and myocardial wall). Also, the proposed SR approach ensures good consistency in the slice direction using no heart shape priors.

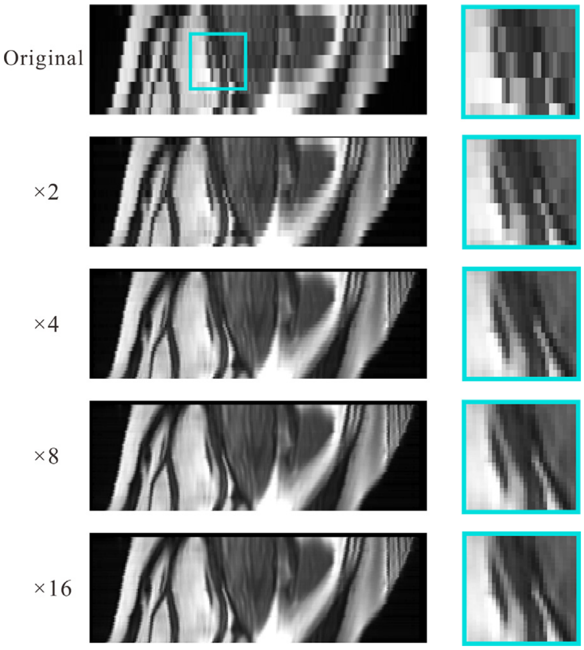

**Fig. 8.** Illustration of progressive improvement of the resolution from long-axis view, where the original, sparsely-sampled cardiac MR image stack was up-sampled with the scale factor of 2, 4, 8 and 16 by recursively employing the proposed SR method.

The visual comparison of an 8x up-sampled, SAX image by the different methods is presented in Fig. 9. Images are shown at both ED and ES phases. As expected, the bicubic interpolation and registration-based interpolation produce blurred images with residual staircasing effects at the border of the ventricles. Although the optical flow interpolation method reasonably restores anatomical structures, results still lack fine details and contain some texture artefacts. On the other hand, the interleaving of the original sharp slices and synthesised, smoothed slices from the depth-aware flow interpolation leads to an inconsistent resolution in the slice direction, which is pronounced in the LAX orientations. The proposed method is superior over all other state-of-the-art interpolation methods in terms of well-preserved structures, edges, image sharpness and consistency. Another patient case shown in Fig. 10 with the comparison of a zoomed-in ROI region highlights the resolution enhancement of the proposed method, such as fine structure restoration of the myocardial wall and papillary muscles.

### 4.2. Segmentation Accuracy

In the second experiment, we evaluated the ventricle segmentation as subsequent image analysis to super-resolved SAX cardiac MR images. First, the intermediate slices synthesised from 10 mm resolution data with different up-sampling methods were automatically segmented to the LV cavity and myocardium using the segmentation method in (Bai et al., 2018). The ground-truth slices available at this resolution were also segmented with the same al-

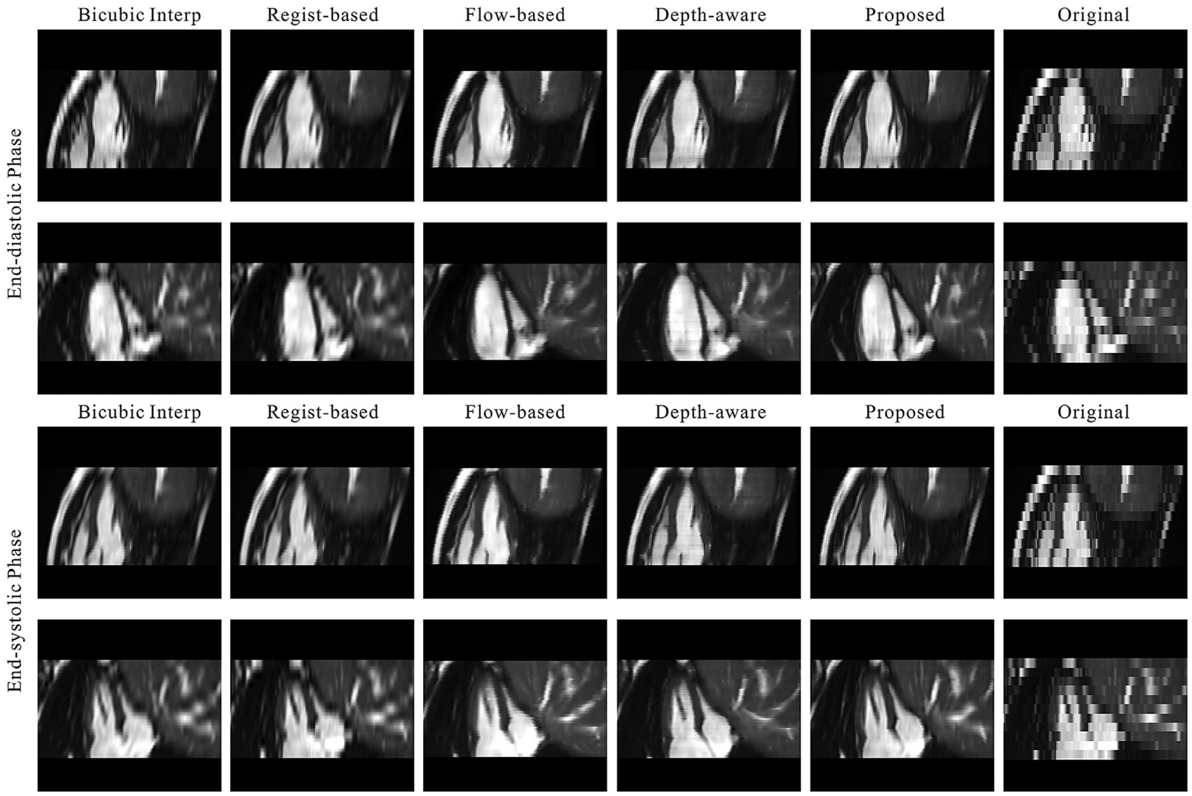

**Fig. 9.** Visual comparison of an 8x up-sampled, SAX image by the different competing methods from long-axis view. Results are shown at both ED and ES phases. From left to right: the linear interpolation, registration-based interpolation, optical flow based frame interpolation, depth-aware flow interpolation, the proposed SR approach and the original, clinically-acquired image.

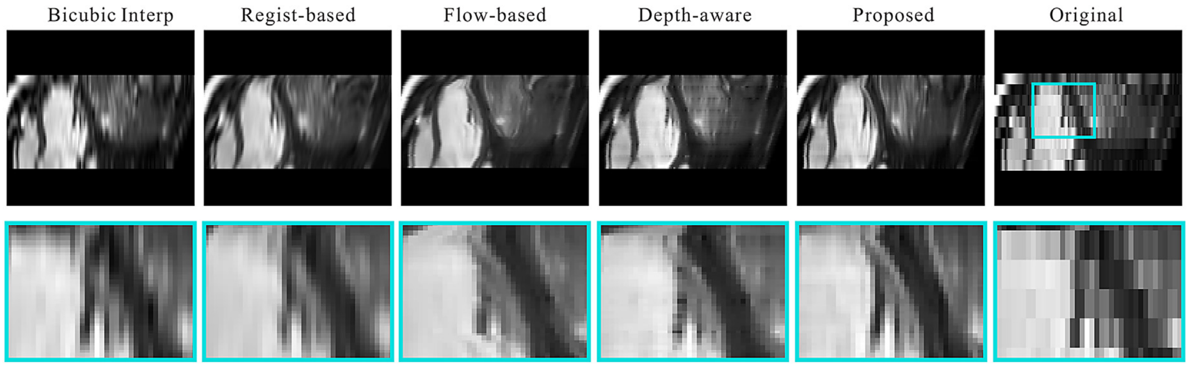

**Fig. 10.** Visual comparison of an 8x up-sampled, SAX image by the different competing methods from long-axis view. From left to right: the linear interpolation, registration-based interpolation, optical flow based frame interpolation, depth-aware flow interpolation, the proposed SR approach and the original, clinically-acquired image. The second row shows the corresponding zoomed-in ROI region, which highlights the resolution enhancement of the proposed approach.

gorithm. Then, the DSC, MCD and HD metrics were computed to assess the segmentation accuracy by comparing those automated segmentation results between ground truth and synthesised intermediate slice. The evaluation was performed on 200 test subjects and shown in Table 1. We can see that our method consistently outperforms the other methods in terms of all metrics. The mean and standard deviations of DSC and MCD for the LV cavity and myocardium are 0.95 ± 0.07, 1.12 ± 1.05 mm, 0.81 ± 0.14, and 1.06 ± 0.75 mm, respectively, indicating better agreement between real and synthesised slice segmentations. Also, it is expected that all metric measures for myocardium are worse than that of the LV cavity, as its annular shape has both endocardial and epicardial edges that may cause equal overlap shifts to produce greater error.

Fig. 11 shows the automated segmentation of the LV and RV on the synthesised slices at different locations (i.e. apical, mid and basal regions), illustrating the excellent quality for automated segmentations on the synthesised slices with the proposed SR method. Images are shown at both ED and ES time frames for side-to-side comparison. Also, to assess the segmentation consistency in the slice direction, automatically delineated epicardial and endocardial contour stacks are shown in Fig. 12, for the original sparsely-sampled cardiac MR image, up-sampled isotropic image from the bicubic interpolation and our proposed method. We observed that there was excellent consistency in the segmentation of the synthesised intermediate slices, using the proposed method, as indicated by the smooth transition at the epicardium and endocardium boundaries.

**Table 1**
Quantitative assessment of the segmentation accuracy by comparing the automated segmented results between the reference slice and synthesised intermediate slices between different competing methods on 200 test sets. The DSC, MCD and HD are used as the metrics. All values are shown as mean ± standard deviation.

| | LV Cavity | | | LV Myocardium | | |
|---|---|---|---|---|---|---|
| | DSC n=200 | MCD [mm] n=200 | HD [mm] n=200 | DSC n=200 | MCD [mm] n=200 | HD [mm] n=200 |
| Linear Interp | 0.915 ± 0.117 | 1.81 ± 2.01 | 5.36 ± 3.45 | 0.742 ± 0.185 | 1.79 ± 2.49 | 7.88 ± 2.97 |
| Regist-based | 0.918 ± 0.101 | 1.73 ± 1.94 | 5.09 ± 2.95 | 0.745 ± 0.188 | 1.68 ± 2.32 | 7.72 ± 2.93 |
| Flow-based | 0.932 ± 0.088 | 1.43 ± 1.24 | 4.54 ± 2.32 | 0.774 ± 0.166 | 1.32 ± 1.16 | 7.14 ± 2.79 |
| Depth-aware | 0.932 ± 0.085 | 1.43 ± 1.20 | 4.52 ± 2.30 | 0.788 ± 0.160 | 1.26 ± 1.05 | 6.98 ± 2.59 |
| Proposed | **0.949 ± 0.068** | **1.12 ± 1.05** | **4.01 ± 1.83** | **0.814 ± 0.141** | **1.06 ± 0.75** | **6.39 ± 2.14** |

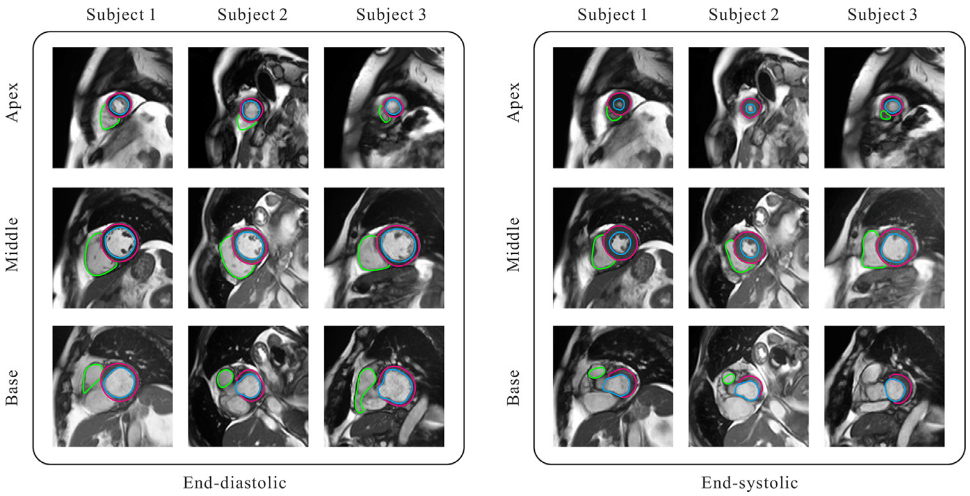

**Fig. 11.** Examples of the segmentation results at the ED and ES phases illustrating the quality for automated segmentations on the synthesised slices with the proposed SR method. Different slice positions from apex to base are chosen and shown. Automated segmentations of LV endocardium, LV myocardium and RV endocardium are delineated in blue, purple and green colour, respectively.

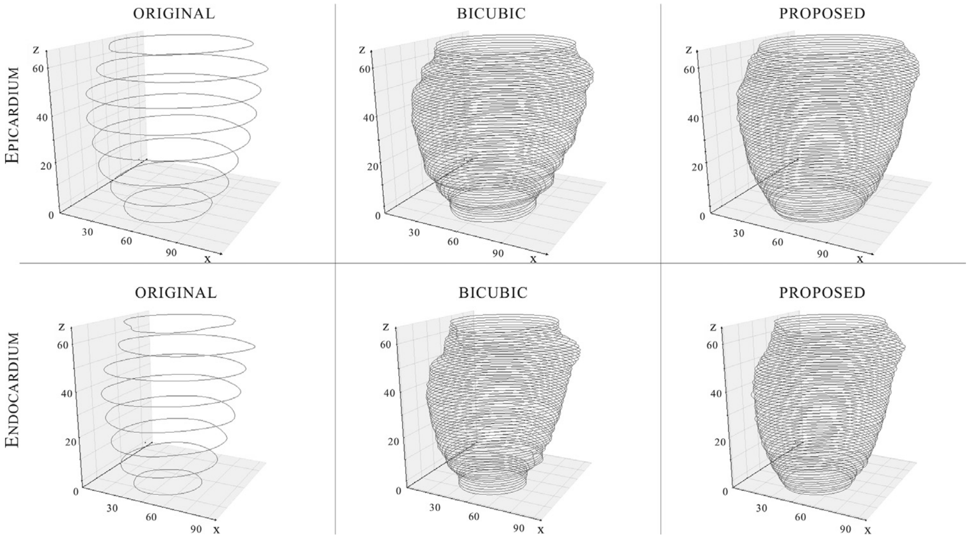

**Fig. 12.** Automatically delineated epicardial and endocardial contour stacks are shown for the original sparsely-sampled cardiac MR image as well as the up-sampled isotropic images by the bicubic interpolation and proposed SR method. The proposed method yields excellent consistency in the segmentation results, highlighted by the smooth, continuous epicardium and endocardium boundaries.

### *4.3. Cardiac Quantification Evaluation*

In this section, we show the impact of our SR method on the accuracy of cardiac measures important to evaluate clinical outcomes. Fig. 13 and Table 2 compare several key clinical parameters including the LV end-diastolic volume (LVEDV) and end-systolic volume (LVESV), LV stroke volume (LVSV), LV ejection fraction (LVEF), LV myocardial mass (LVM), RV end-diastolic volume (RVEDV) and end-systolic volume (RVESV), RV stroke volume (RVSV) and RV ejection fraction (RVEF) computed from 400 automated segmentation results, between the original resolution image stacks (10 mm), down-sampled image stacks (20 mm), three levels of up-sampled image stacks by the proposed method (5 mm, 2.5 mm and 1.25 mm). As illustrated in Fig. 13, we can see that the traditional Simpson's rule of volume approximation differs regarding the input image through-plane resolutions, but the pro-

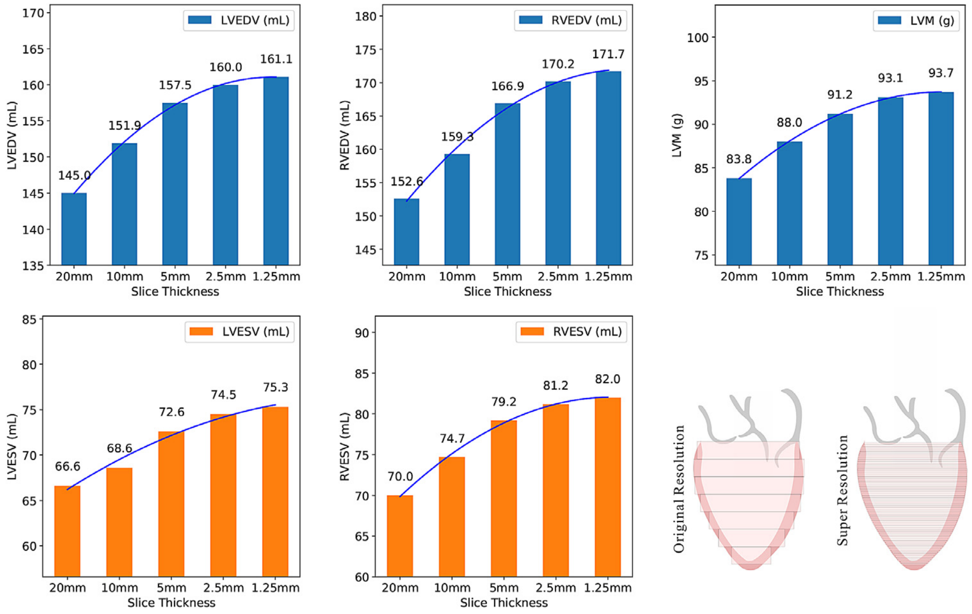

**Fig. 13.** Illustration of the variation in clinical measures (i.e., LVEDV, LVESV, LVM, RVEDV and RVESV) derived from automated segmentation of 400 cardiac MR image stacks with different through-plane resolutions, demonstrating more accurate and stable volume calculation obtained by the high-resolution, isotropic image stacks.

**Table 2**

Mean and standard deviation of the clinical measures derived from 400 automated segmentation results, between the different resolution images. LVEDV and LVESV represent LV end-diastolic volume and end-systolic volume, LVSV and LVEF represent LV stroke volume and ejection fraction, LVM represents LV myocardial mass, RVEDV and RVESV represent RV end-diastolic volume and end-systolic volume, RVSV and LVEF represent RV stroke volume and ejection fraction.

| Slice Resolution | 20 mm ($R_1$) | 10 mm ($R_2$) | 5 mm ($R_3$) | 2.5 mm ($R_4$) | 1.25 mm ($R_5$) | $R_1$ vs $R_2$ | $R_2$ vs $R_3$ | $R_3$ vs $R_4$ | $R_4$ vs $R_5$ |
|---|---|---|---|---|---|---|---|---|---|
| | Mean ± Std. (n=400) | Mean ± Std. (n=400) | Mean ± Std. (n=400) | Mean ± Std. (n=400) | Mean ± Std. (n=400) | p-value | p-value | p-value | p-value |
| LVEDV (mL) | 145.0 ± 38.2 | 151.9 ± 36.5 | 157.5 ± 37.6 | 160.0 ± 38.1 | 161.1 ± 38.3 | <0.05 | <0.05 | 0.28 | **0.63** |
| LVESV (mL) | 66.6 ± 25.0 | 68.6 ± 24.3 | 72.6 ± 25.1 | 74.5 ± 25.5 | 75.3 ± 25.7 | 0.17 | <0.05 | 0.23 | **0.61** |
| LVSV (mL) | 78.4 ± 22.1 | 83.3 ± 19.0 | 84.9 ± 18.9 | 85.5 ± 19.2 | 85.8 ± 19.4 | <0.05 | 0.16 | 0.258 | **0.77** |
| LVEF (%) | 54.5 ± 9.0 | 55.4 ± 7.0 | 54.5 ± 6.5 | 54.0 ± 6.5 | 53.8 ± 6.5 | 0.05 | <0.05 | 0.28 | **0.67** |
| LVM (g) | 83.8 ± 24.5 | 88.0 ± 24.6 | 91.2 ± 25.2 | 93.1 ± 25.8 | 93.7 ± 26.0 | <0.05 | <0.05 | 0.22 | **0.70** |
| RVEDV (mL) | 152.6 ± 40.8 | 159.3 ± 38.8 | 166.9 ± 40.1 | 170.2 ± 40.6 | 171.7 ± 40.9 | <0.05 | <0.05 | 0.17 | **0.55** |
| RVESV (mL) | 70.0 ± 24.8 | 74.7 ± 23.6 | 79.2 ± 24.5 | 81.2 ± 25.0 | 82.0 ± 25.2 | <0.05 | <0.05 | 0.178 | **0.59** |
| RVSV (mL) | 82.5 ± 22.8 | 84.5 ± 21.1 | 87.8 ± 21.5 | 89.0 ± 21.8 | 89.7 ± 22.0 | 0.12 | <0.05 | 0.32 | **0.63** |
| RVEF (%) | 54.6 ± 8.0 | 53.5 ± 6.9 | 53.0 ± 6.8 | 52.7 ± 6.8 | 52.6 ± 6.8 | <0.05 | 0.21 | 0.51 | **0.86** |

posed SR method reduces the variability of ventricular volumetric assessment by providing high-resolution, 3D isotropic image stacks that enable more accurate and reliable volume calculations; see the small variations between resolutions of 2.5 mm and 1.25 mm in LVEDV, LVESV, LVM, RVEDV and RVESV, which are reflected by the gentle slopes of the mean value curves in Fig. 13. For image volumes with through-plane resolution higher than 5 mm, we did not observe any significant change in the derived morphological and functional indices. When the through-plane resolution is higher than 1.25 mm, the derived morphological and functional indices are primarily determined by the in-plane resolution and segmentation accuracy. Statistical significance of the results was verified by performing the Student's t-test between the derived cardiac indices from different resolutions (cf. Table 2). It is worth mentioning that there were statistically significant differences in all cardiac indices, when comparing the two lowest resolutions (20 mm and 10 mm) with respect to the two highest resolutions (2.5 mm and 1.25 mm).

We also evaluated the myocardial wall thickness at the cardiac ED and ES phases. A comparison of wall thickness maps derived from manual delineations and automated segmentations with different methods is shown in Fig. 14. Wall thickness was computed by estimating the minimum distance between epicardium and endocardium in the subject space, and then mapped and represented onto an atlas surface mesh provided in (Bai et al., 2015). We can see that the SR method facilitated a smooth, continuous distribution in the analysis of myocardial wall thickness, which yielded more anatomically plausible and precise results. Fig. 15 shows the bulls-eye plots of the regional wall thickness analysis in 400 test subjects based on the AHA 17-segment model. We observed that the SR method provided better agreement with the results derived from the manual delineations, particularly near the apical (regions 13-17) and basal regions (regions 1-6), where the ventricular contours are more complex and thus more difficult to segment.

### *4.4. Non-rigid Registration Accuracy*

3D non-rigid registration was tested on 100 cardiac MR cine images within the full cardiac cycle using 50 time points (i.e., 5,000 volumes for each method) for the original clinically-acquired images and super-resolved, 3D isotropic images to verify the accuracy of the HR registration compared to the native LR registration.

Fig. 16 shows the displacement field (overlaid on the SAX slices) at the forward pass (ED to ES) and backward pass (ES to ED) for both the LV and RV. With HR registration, the displacement vectors (represented by arrows) clearly indicate that during the contraction the ventricles are moving inwards, whilst they are moving outwards at the period of relaxation.

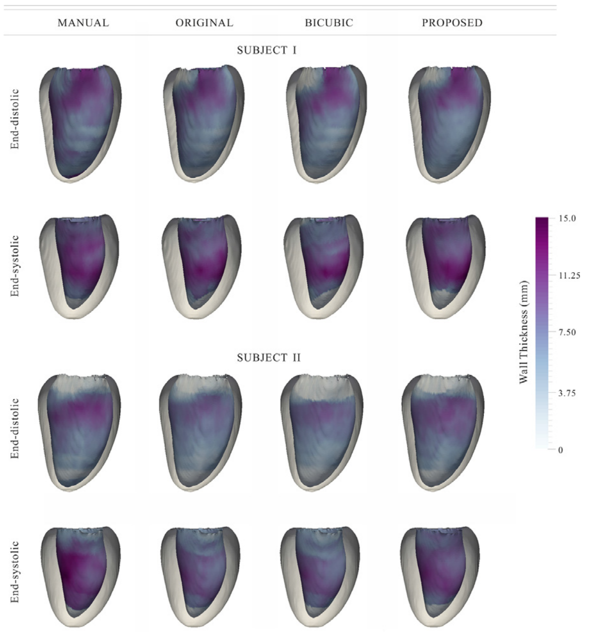

**Fig. 14.** Myocardial wall thickness map derived from manual delineations on the original image, automated segmentation on the original image and the up-sampled isotropic images by the bicubic interpolation and proposed SR method. Wall thickness was obtained by estimating the minimum distance between epicardium and endocardium in the subject space and then mapped onto a template surface mesh for visual comparison.

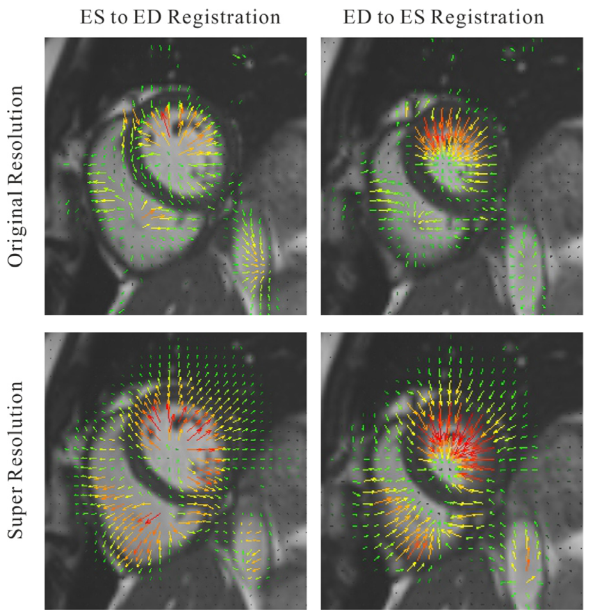

**Fig. 16.** Visualisation of the displacement field obtained in the frame-to-frame registration from ED to ES phase and from ES to ED phase.

We quantitatively evaluated the performance of up-sampled, isotropic image stacks for the non-rigid registration in the cardiac sequence. As described in Section 3.4, we warped segmentation masks at different cardiac phases $t \in [1, 49]$ to the ED phase $t$=0 using the computed frame-to-reference transformation $\varphi_{v_t \rightarrow v_0}$ and then estimated the DSC, MCD and HD metrics between the ED segmentation and warped segmentation. Fig. 17 shows the mean and standard deviation of all metrics in these 49 cardiac phases for the 100 subjects. As expected, we found an error accumulation during the cardiac contraction until the ES phase for the frame to ED reference registration. Both the up-sampling methods outperform the native LR registration by a large margin in most time points, particularly during the maximum contraction. Also, further improvement using the proposed method is demonstrated by the mean and standard deviation of DSC, MCD and HD over all time points: 0.86 ± 0.03 vs. 0.83 ± 0.03, 0.99 ± 0.18 mm vs. 1.24 ± 0.21 mm, and 2.46 ± 0.28 mm vs. 2.87 ± 0.37 mm by comparing with the bicubic interpolation for myocardium registration, respectively.

### 4.5. Ablation Study

This section presents the ablation study results assessing the contribution of the adopted components in the generative model. Three variants, namely w/o CBN, w/o CBN + FM and w/o CBN + FM + MD, corresponding to replacing the CBN layers with standard batch normalisation, replacing feature matching loss with pixel-wise loss and using a single discriminator for the proposed model respectively, were investigated and compared with the standard full version. For fair comparison, we retrained these variant net-

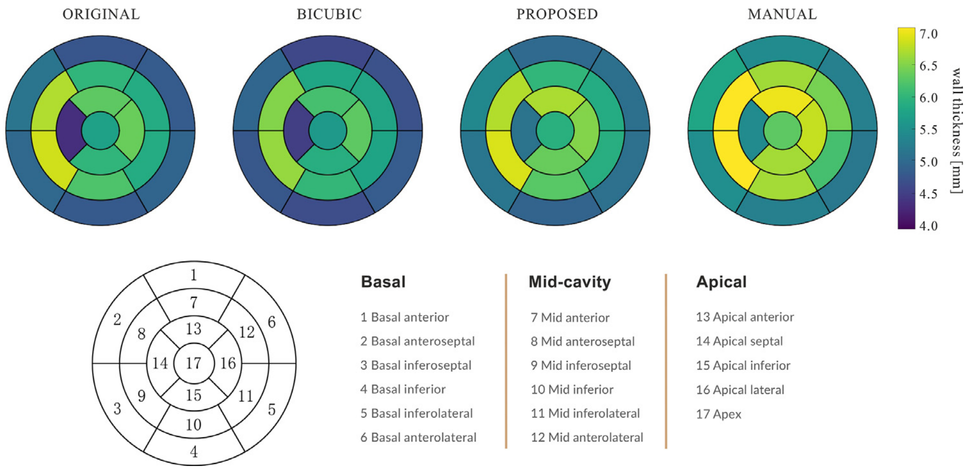

**Fig. 15.** Bulls-eye plots of the regional wall thickness analysis (AHA 17-segment model) obtained by averaging results of 400 subjects. The proposed method was compared with manual delineations on the original images, automated segmentations on the original images as well as the up-sampled isotropic images by the bicubic interpolation.

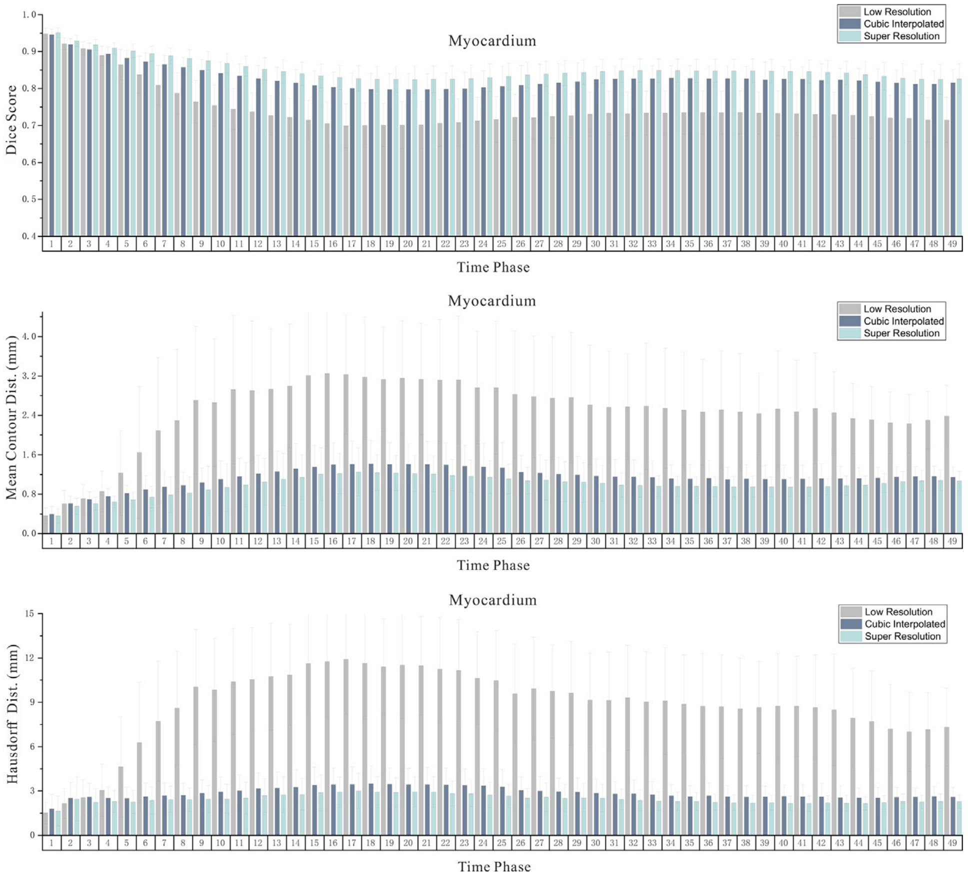

**Fig. 17.** Summary of the frame-to-reference registration error throughout 49 time points in the cardiac sequence on 100 test sets. The results of the super-resolved images from the proposed method was compared with those of the original LR images and the up-sampled isotropic volumes by the bicubic interpolation. DSC, MCD and HD are used as the metrics.

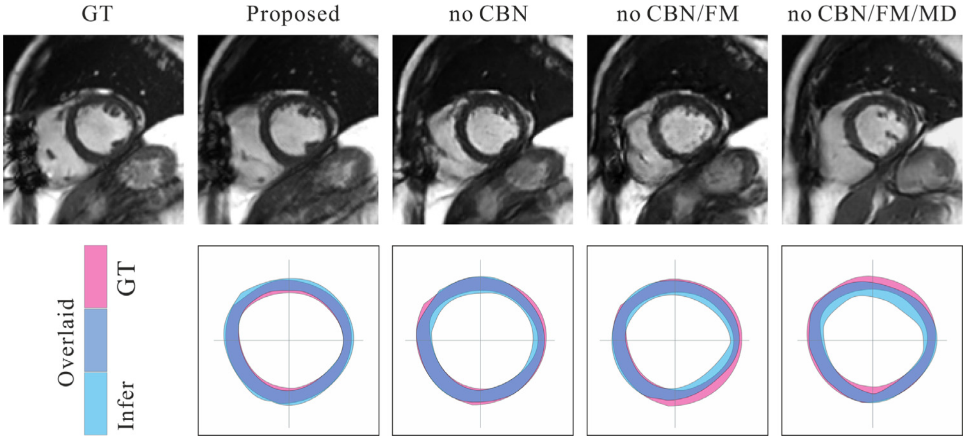

**Fig. 18.** Visual inspection of the ablation study results on the generated images. The second row shows the corresponding myocardial segmentation of each variant (shown in light blue colour) overlaid with the myocardial segmentation of the ground-truth image (shown in pink colour).

works using the same hyperparameters and training epochs as the proposed method.

Fig. 18 shows the visual inspection of the ablation study on the generated images and the corresponding myocardial segmentations. The proposed method with all the components integrated yields results not only visually comparable to that of the reference, but also can retain the accuracy of myocardial segmentation that many subsequent image analyses rely on. Table 3 shows the quantitative results of the ablation study, indicating that the CBN and discriminator-based feature matching loss produce better results than the conventional batch normalisation and pixel-wise loss term. Removing or replacing them results in a considerable performance decrease (e.g., a PSNR of $21.53 \pm 1.78$ to $19.76 \pm 1.19$ for image synthesis). We can see that the multi-scale discriminator further boosts to the synthesis performance, which is reflected by an average reduction of 2.68 mm and 2.36 mm in Hausdorff distance for the LV cavity and myocardium segmentation, respectively.

## 5. Discussions

In this work, we aim to address the problem of generating HR images from single LR cardiac MR stacks acquired in typical clin-

**Table 3**
Summary of the ablation study results evaluated on the generated images and the corresponding automated segmentations on 200 test sets.

| Metrics | Generated Image | | Segmentation (Endo) | | Segmentation (Myo) | |
|---|---|---|---|---|---|---|
| | CC n=200 | PSNR [dB] n=200 | DSC n=200 | HD [mm] n=200 | DSC n=200 | HD [mm] n=200 |
| Proposed | **0.884 ± 0.058** | **21.53 ± 1.78** | **0.949 ± 0.068** | **4.01 ± 1.83** | **0.814 ± 0.140** | **6.39 ± 2.14** |
| w/o CBN | 0.847 ± 0.038 | 19.96 ± 1.15 | 0.899 ± 0.128 | 5.06 ± 2.29 | 0.718 ± 0.199 | 7.44 ± 2.20 |
| w/o CBN+FM | 0.825 ± 0.043 | 19.76 ± 1.19 | 0.893 ± 0.126 | 5.63 ± 2.79 | 0.659 ± 0.190 | 8.52 ± 2.49 |
| w/o CBN+FM+MD | 0.843 ± 0.035 | 19.34 ± 0.98 | 0.880 ± 0.149 | 8.31 ± 6.76 | 0.591 ± 0.183 | 10.88 ± 4.27 |

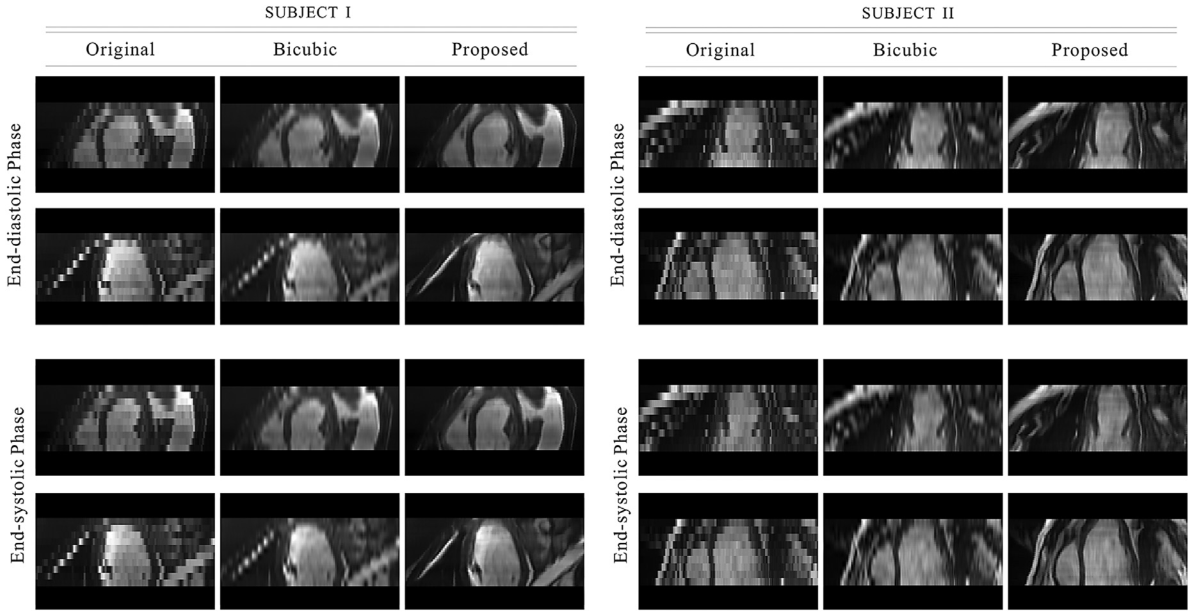

**Fig. 19.** Visual assessment of the proposed generative model trained on the UKBB dataset and tested on the ACDC dataset. Two subject examples are shown at both ED and ES phases.

ical settings that is not yet completely solved to the best of our knowledge. Typical SR techniques require either paired HR and LR images or multiple 2D LR stacks from different orientations, often not available in real-world clinical settings. Natural video frame interpolation methods that can synthesise an intermediate frame for unseen data do not perform well in this scenario since the slice sampling in cardiac MR image stacks is much lower than temporal sampling of neural video frames, and visual differences and quantitative information in medical images are more complex and non-trivial than natural scene videos. For these reasons, up-sampling of cardiac images via intensity-based interpolation schemes, such as B-spline or bicubic, is still most frequently used in cardiac image analysis, although several artefacts remain such as blurring and loss in soft tissue contrast, among others. To tackle these issues, we proposed a novel, robust adversarial learning algorithm based on conditional GAN and employed a state-of-the-art optical flow method to generate an auxiliary image to train the network end-to-end in an unsupervised transfer learning fashion. The proposed framework effectively incorporates visual properties and relevant textures of input images and can synthesise HR anatomically plausible cardiac MR images consistent with the available slices. To synthesise visually appealing images and facilitate the estimation of accurate quantitative measurements, we employed a dedicated generator, discriminator and optimisation procedure that led to a better SR model in this work.

First, rather than using a deeper network that would increase the network capacity and cause overfitting, a multi-scale discriminator was employed to operate on different receptive fields that could simultaneously encourage the generator to synthesise globally coherent images and capture finer structural details. Second, replacing the traditional pixel-wise loss functions with feature matching loss facilitated better recovery of high-frequency image information and stabilised the adversarial training. Finally, several studies have demonstrated that CBN-based generators are highly effective in a wide variety of contexts (Ulyanov et al., 2017; Huang and Belongie, 2017; Chen et al., 2018a; Park et al., 2019). The mode of action behind CBNs is to incorporate the additional auxiliary data or guidance image into the conditional generative model through normalisation layers, to guide the image generation conditioned on certain constraints. In our case, it is intuitive to use two adjacent slices as guidance to capture spatial features from regions near the synthesised, intermediate slice position. The ablation study in Section 4.5 showed these architectural adaptations help boost overall model performance.

Through comprehensive experiments, we have shown that our approach outperformed the state-of-the-art baselines both qualitatively and quantitatively. Also, we demonstrated the added value of super-resolved images on subsequent image analysis, including ventricle segmentation, cardiac quantification and non-rigid registration. Intuitively, isotropic cardiac MR images would be useful for understanding complex anatomy and function in heart diseases and visualisation and analysis of small cardiac structures such as the atrium or the valves. It also helps to reduce the variability of ventricular volumetric assessment, which is generally high with the 2D image stacks due to 1) the variation in the conventional volume calculation (e.g. via the Simpson's rule) from low through-plane resolution and 2) the difficulty of automatically segmenting the myocardium near the apical and basal regions of the ventricles

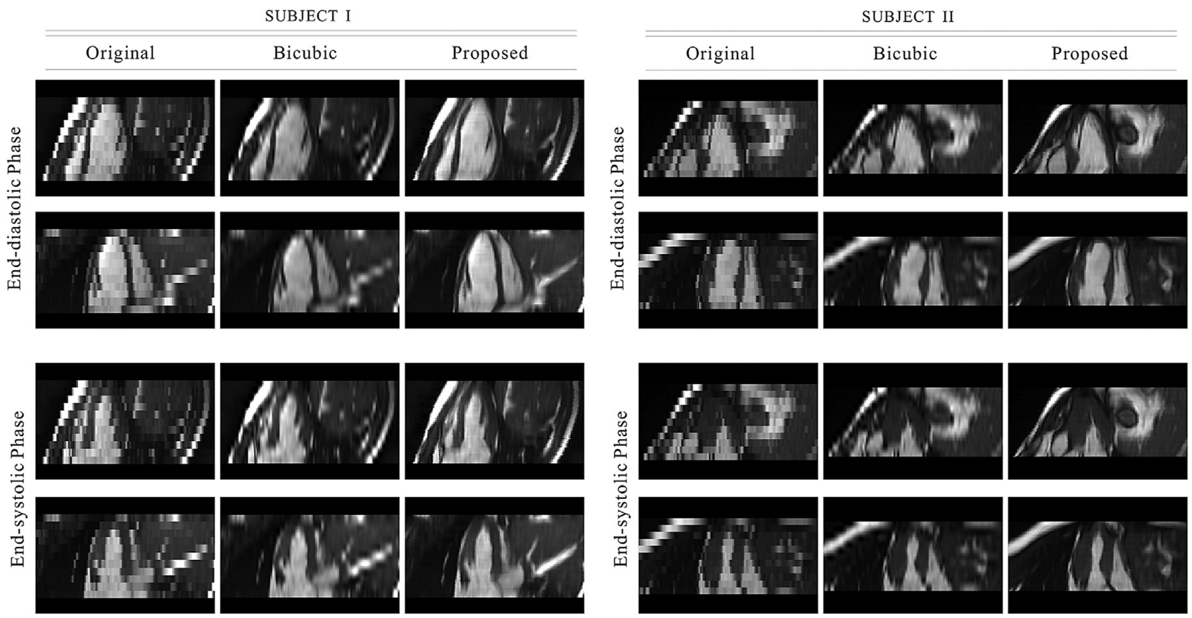

**Fig. 20.** Visual assessment of the proposed generative model trained on the UKBB dataset and tested on the ADSB dataset. Two subject examples are shown at both ED and ES phases.

in LR images, as illustrated in Figs. 13 and 15. Also, conventional data augmentation for training segmentation models only involves increasing training samples by using simple transformations such as rotation, flipping, scaling etc. With the proposed SR algorithm, one can augment the training data by generating intermediate images and the corresponding segmentation labels.

Another aspect is that the highly non-rigid heart motion present in cardiac imaging, with the low through-plane resolution, which makes accurate non-rigid registration challenging to achieve in cardiac cine images. Many studies suggested that the performance of non-rigid registration may differ regarding the input image resolution, and the ratio between in-plane and through-plane resolutions (ur Rahman and Wesarg, 2010; Van Reeth et al., 2012). The proposed method enabled the generation of isotropic image stacks and facilitates HR cardiac image registration and analysis in a sub-pixel accuracy, compared with original LR images. As shown in Fig. 17, the SR method yielded more stable and accurate frame-to-reference registration results, particularly for the time frames far away from the ED phase. We further demonstrated that the proposed method outperformed the traditional bicubic interpolation, which is commonly utilised in practice.

Although analysing heart motion provides crucial insights to cardiac function, estimating myocardial strain in cardiac cine MR imaging is still challenging: 1) The diffeomorphic image registration inability to accurately estimate the heart deformation in LR images; 2) Low through-plane resolution leads to an underestimation of longitudinal strains; 3) Highly anisotropic volumes may cause inaccurate ventricle/myocardium parcellation that negatively influences motion data dimensionality reduction. It would be interesting to systemically investigate how the SR method improves 3D strain analysis in future work.

The optical flow method used to generate the auxiliary image in this work was initially developed for interpolating intermediate frames for large object motion in dynamic sequences. This is beneficial for cardiac cine MR imaging as the method can still fill in in-between slices that lie precisely between and central to the neighbouring slices, even in the presence of slice misalignment and inter-slice motion artefacts. However, the proposed method cannot eliminate such slice misalignment and inter-slice motion artefacts. To address this problem, instead of explicitly employing a quality control or motion compensation algorithm, it would also be interesting to incorporate additional physical constraints or shape priors into the generative model, which may help synthesise more reliable, shape-consistent images for analyses and visualisation.

## 6. Conclusion

In this work, we proposed a conditional GAN-based method for generating HR, 3D isotropic cardiac MR cine images. The approach considers the real clinical scenario and requires neither the corresponding HR scans nor multiple LR scans with different orientations for training. The proposed approach can be trained in an unsupervised transfer learning manner using an adversarial learning strategy, to synthesise anatomically plausible images consistent with the available slices. Experimental results showed that the proposed SR method outperformed traditional interpolation methods and video frame interpolations qualitatively and quantitatively. Importantly, we demonstrated that subsequent image analyses such as ventricle segmentation, cardiac quantification and non-rigid registration benefit from the proposed SR method by generating super-resolved, 3D isotropic cardiac MR images, producing more accurate quantitative results, without requiring a long acquisition time. Although in this work we only addressed the SR problem in cardiac MR images, our approach is generic and could be applied to other anatomical regions or modalities, promising to facilitate more accurate clinical analyses.

## Declaration of Competing Interest

Steffen E. Petersen provides consultancy to and is shareholder of Circle Cardiovascular Imaging Inc, Calgary, Canada.

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## CRediT authorship contribution statement

**Yan Xia:** Conceptualization, Methodology, Software, Visualization, Validation, Formal analysis, Writing - original draft. **Nishant Ravikumar:** Conceptualization, Methodology, Writing - review &

editing. **John P. Greenwood:** Writing - review & editing. **Stefan Neubauer:** Data curation, Writing - review & editing. **Steffen E. Petersen:** Data curation, Writing - review & editing. **Alejandro F. Frangi:** Conceptualization, Supervision, Writing - review & editing.

## Acknowledgements

This research has been conducted using the UK Biobank Resource under Application 11350. The cardiac MR images presented in Figs. 1, 2, 6, 7, 8, 9, 10, 11, 16 and 18 in the manuscript were reproduced with the permission of UK Biobank ©. The authors are grateful to all UK Biobank participants and staff. AFF acknowledges support from the Royal Academy of Engineering Chair in Emerging Technologies Scheme (CiET1819/19), EPSRC-funded Grow MedTech CardioX (POC041), and the MedIAN Network (EP/N026993/1) funded by the Engineering and Physical Sciences Research Council (EPSRC). The work of AFF is also partially funded by EPSRC through TUSCA (EP/V04799X/1). SN acknowledges the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre based at The Oxford University Hospitals Trust at the University of Oxford, and the British Heart Foundation Centre of Research Excellence. SEP acknowledges support from the SmartHeart EPSRC programme grant (www.nihr.ac.uk; EP/P001009/1), the British Heart Foundation for funding the manual analysis to create a cardiovascular magnetic resonance imaging reference standard for the UK Biobank imaging resource in 5,000 CMR scans (www.bhf.org.uk; PG/14/89/31194), and the National Institute for Health Research (NIHR) Barts Biomedical Research Centre. SEP has received funding from the European Union' s Horizon 2020 research and innovation programme under grant agreement No 825903 (euCanSHare project).

## Appendix A

We tested the trained generative model on two additional, publicly accessible cardiac MR cine datasets: Automatic Cardiac Diagnosis Challenge (ACDC) dataset [2]and Kaggle Second Annual Data Science Bowl (ADSB) dataset [3]. Cine MR images of the ACDC dataset were acquired with a conventional SSFP sequence from two MRI scanners of different magnetic strengths (1.5 T - Siemens Area, Siemens Medical Solutions, Germany and 3.0 T - Siemens Trio Tim, Siemens Medical Solutions, Germany), with slice thickness ranging from 5 mm to 10 mm. The in-plane spatial resolution varies from 1.34 to 1.68 mm. Each sequence contains 28-40 cardiac phases covering completely or partially one cardiac cycle. The ADSB dataset was compiled by the National Institutes of Health and Children's National Medical Center. The slice thickness ranges from 6 mm to 8 mm and the in-plane spatial resolution varies from 0.61 to 1.75 mm. Each subject contains 30 cardiac phases and the number of slices in the SAX stack typically ranges between 8 and 14. Qualitative results of these two additional datasets are shown in Figs. 19 and 20.

## References

Anand, S.S., Tu, J.V., Awadalla, P., Black, S., Boileau, C., Busseuil, D., Desai, D., Després, J.-P., de Souza, R.J., Dummer, T., et al., 2016. Rationale, design, and methods for Canadian alliance for healthy hearts and minds cohort study (CAHHM)–a Pan Canadian cohort study. BMC public health 16 (1), 650.

Bai, W., Shi, W., de Marvao, A., Dawes, T.J.W., O', D.P., Cook, S.A., Rueckert, D., 2015. A bi-ventricular cardiac atlas built from 1000+ high resolution MR images of healthy subjects and an analysis of shape and motion. Medical image analysis 26 (1), 133–145.

Bai, W., Sinclair, M., Tarroni, G., Oktay, O., Rajchl, M., Vaillant, G., Lee, A.M., Aung, N., Lukaschuk, E., Sanghvi, M.M., et al., 2018. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. Journal of Cardiovascular Magnetic Resonance 20 (1), 65.

Bamberg, F., Kauczor, H.-U., Weckbach, S., Schlett, C.L., Forsting, M., Ladd, S.C., Greiser, K.H., Weber, M.-A., Schulz-Menger, J., Niendorf, T., et al., 2015. Whole-body MR imaging in the German National Cohort: rationale, design, and technical background. Radiology 277 (1), 206–220.

Bao, W., Lai, W.-S., Ma, C., Zhang, X., Gao, Z., Yang, M.-H., 2019. Depth-aware video frame interpolation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3703–3712.

Bao, W., Lai, W.-S., Zhang, X., Gao, Z., Yang, M.-H., 2019. Memc-net: Motion estimation and motion compensation driven neural network for video interpolation and enhancement. IEEE transactions on pattern analysis and machine intelligence.

Basty, N., Grau, V., 2018. Super resolution of cardiac cine MRI sequences using deep learning. In: Image Analysis for Moving Organ, Breast, and Thoracic Images. Springer, pp. 23–31.

Chaudhari, A.S., Fang, Z., Kogan, F., Wood, J., Stevens, K.J., Gibbons, E.K., Lee, J.H., Gold, G.E., Hargreaves, B.A., 2018. Super-resolution musculoskeletal MRI using deep learning. Magnetic resonance in medicine 80 (5), 2139–2154.

Chen, T., Lucic, M., Houlsby, N., Gelly, S., 2018. On self modulation for generative adversarial networks. arXiv preprint arXiv:1810.01365.

Chen, Y., Shi, F., Christodoulou, A.G., Xie, Y., Zhou, Z., Li, D., 2018. Efficient and accurate MRI super-resolution using a generative adversarial network and 3D multi--level densely connected network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 91–99.

Dalca, A.V., Bouman, K.L., Freeman, W.T., Rost, N.S., Sabuncu, M.R., Golland, P., 2018. Medical image imputation from image collections. IEEE transactions on medical imaging 38 (2), 504–514.

De Vries, H., Strub, F., Mary, J., Larochelle, H., Pietquin, O., Courville, A.C., 2017. Modulating early visual processing by language. In: Advances in Neural Information Processing Systems, pp. 6594–6604.

Dong, C., Loy, C.C., He, K., Tang, X., 2014. Learning a deep convolutional network for image super-resolution. In: European conference on computer vision. Springer, pp. 184–199.

Durugkar, I., Gemp, I., Mahadevan, S., 2017. Generative multi-adversarial networks. In: International Conference on Learning Representations (ICLR).

Frakes, D.H., Dasi, L.P., Pekkan, K., Kitajima, H.D., Sundareswaran, K., Yoganathan, A.P., Smith, M.J.T., 2008. A new method for registration-based medical image interpolation. IEEE transactions on medical imaging 27 (3), 370–377.

Gholipour, A., Estroff, J.A., Warfield, S.K., 2010. Robust super-resolution volume reconstruction from slice acquisitions: application to fetal brain MRI. IEEE transactions on medical imaging 29 (10), 1739–1758.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. In: Advances in neural information processing systems, pp. 2672–2680.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.

Herbst, E., Seitz, S., Baker, S., 2009. Occlusion reasoning for temporal interpolation using optical flow. Department of Computer Science and Engineering, University of Washington, Tech. Rep. UW-CSE-09-08-01.

Horváth, A., Pezold, S., Weigel, M., Parmar, K., Cattin, P., 2017. High order slice interpolation for medical images. In: International Workshop on Simulation and Synthesis in Medical Imaging. Springer, pp. 69–78.

Huang, X., Belongie, S., 2017. Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1501–1510.

Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1125–1134.

Jiang, H., Sun, D., Jampani, V., Yang, M.-H., Learned-Miller, E., Kautz, J., 2018. Super slomo: High quality estimation of multiple intermediate frames for video interpolation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9000–9008.

Jog, A., Carass, A., Prince, J.L., 2014. Improving magnetic resonance resolution with supervised learning. In: 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI). IEEE, pp. 987–990.

Karras, T., Aila, T., Laine, S., Lehtinen, J., 2018. Progressive growing of GANs for improved quality, stability, and variation. International Conference on Learning Representations (ICLR).

Knauth, A.L., Gauvreau, K., Powell, A.J., Landzberg, M.J., Walsh, E.P., Lock, J.E., del Nido, P.J., Geva, T., 2008. Ventricular size and function assessed by cardiac MRI predict major adverse clinical outcomes late after tetralogy of Fallot repair. Heart 94 (2), 211–216.

Konukoglu, E., van der Kouwe, A., Sabuncu, M.R., Fischl, B., 2013. Example-based restoration of high-resolution magnetic resonance image acquisitions. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 131–138.

Kudo, A., Kitamura, Y., Li, Y., Iizuka, S., Simo-Serra, E., 2019. Virtual thin slice: 3D conditional GAN-based Super-resolution for CT slice interval. In: International Workshop on Machine Learning for Medical Image Reconstruction. Springer, pp. 91–100.

Leng, J., Xu, G., Zhang, Y., 2013. Medical image interpolation based on multi-resolution registration. Computers & Mathematics with Applications 66 (1), 1–18.

Liu, Z., Yeh, R.A., Tang, X., Liu, Y., Agarwala, A., 2017. Video frame synthesis using

---

[2] https://www.creatis.insa-lyon.fr/Challenge/acdc/

[3] https://www.kaggle.com/c/second-annual-data-science-bowl/

deep voxel flow. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 4463–4471.

Lyu, Q., You, C., Shan, H., Zhang, Y., Wang, G., 2019. Super-resolution MRI and CT through GAN-circle. In: Developments in X-Ray Tomography XII, Vol. 11113. International Society for Optics and Photonics, p. 111130X.

Manjón, J.V., Coupé, P., Buades, A., Collins, D.L., Robles, M., 2010. MRI superresolution using self-similarity and image priors. International journal of biomedical imaging 2010.

Manjón, J.V., Coupé, P., Buades, A., Fonov, V., Collins, D.L., Robles, M., 2010. Non-local MRI upsampling. Medical image analysis 14 (6), 784–792.

Mao, X., Li, Q., Xie, H., Lau, R.Y.K., Wang, Z., Paul Smolley, S., 2017. Least squares generative adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2794–2802.

Marwick, T.H., Neubauer, S., Petersen, S.E., 2013. Use of cardiac magnetic resonance and echocardiography in population-based studies: why, where, and when? Circulation: Cardiovascular Imaging 6 (4), 590–596.

Mirza, M., Osindero, S., 2014. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.

Miyato, T., Kataoka, T., Koyama, M., Yoshida, Y., 2018. Spectral normalization for generative adversarial networks. International Conference on Learning Representations (ICLR).

Miyato, T., Koyama, M., 2018. cGANs with projection discriminator. arXiv preprint arXiv:1802.05637.

Nguyen, T., Le, T., Vu, H., Phung, D., 2017. Dual discriminator generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2670–2680.

Niklaus, S., Liu, F., 2018. Context-aware synthesis for video frame interpolation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1701–1710.

Oktay, O., Bai, W., Lee, M., Guerrero, R., Kamnitsas, K., Caballero, J., de Marvao, A., Cook, S., O', D., Rueckert, D., 2016. Multi-input cardiac image super-resolution using convolutional neural networks. In: International conference on medical image computing and computer-assisted intervention. Springer, pp. 246–254.

Park, T., Liu, M.-Y., Wang, T.-C., Zhu, J.-Y., 2019. Semantic image synthesis with spatially-adaptive normalization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2337–2346.

Petersen, S.E., Aung, N., Sanghvi, M.M., Zemrak, F., Fung, K., Paiva, J.M., Francis, J.M., Khanji, M.Y., Lukaschuk, E., Lee, A.M., et al., 2017. Reference ranges for cardiac structure and function using cardiovascular magnetic resonance (CMR) in Caucasians from the UK Biobank population cohort. Journal of Cardiovascular Magnetic Resonance 19 (1), 18.

Petersen, S.E., Matthews, P.M., Bamberg, F., Bluemke, D.A., Francis, J.M., Friedrich, M.G., Leeson, P., Nagel, E., Plein, S., Rademakers, F.E., et al., 2013. Imaging in population science: cardiovascular magnetic resonance in 100,000 participants of UK Biobank-rationale, challenges and approaches. Journal of Cardiovascular Magnetic Resonance 15 (1), 46.

Petersen, S.E., Matthews, P.M., Francis, J.M., Robson, M.D., Zemrak, F., Boubertakh, R., Young, A.A., Hudson, S., Weale, P., Garratt, S., et al., 2015. Uk biobanks cardiovascular magnetic resonance protocol. Journal of cardiovascular magnetic resonance 18 (1), 8.

Plenge, E., Poot, D.H.J., Niessen, W.J., Meijering, E., 2013. Super-resolution reconstruction using cross-scale self-similarity in multi-slice MRI. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 123–130.

ur Rahman, S., Wesarg, S., 2010. Upsampling of cardiac MR images: Comparison of averaging and super-resolution for the combination of multiple views. In: Proceedings of the 10th IEEE International Conference on Information Technology and Applications in Biomedicine. IEEE, pp. 1–4.

Rousseau, F., Habas, P.A., Studholme, C., 2011. A supervised patch-based approach for human brain labeling. IEEE transactions on medical imaging 30 (10), 1852–1862.

Scheffler, K., Lehnhardt, S., 2003. Principles and applications of balanced SSFP techniques. European radiology 13 (11), 2409–2418.

Ulyanov, D., Vedaldi, A., Lempitsky, V., 2017. Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6924–6932.

Van Reeth, E., Tham, I.W.K., Tan, C.H., Poh, C.L., 2012. Super-resolution in magnetic resonance imaging: a review. Concepts in Magnetic Resonance Part A 40 (6), 306–325.

Vercauteren, T., Pennec, X., Perchant, A., Ayache, N., 2007. Non-parametric diffeomorphic image registration with the demons algorithm. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 319–326.

Wang, J., Chen, Y., Wu, Y., Shi, J., Gee, J., 2020. Enhanced generative adversarial network for 3D brain MRI super-resolution. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 3627–3636.

Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., Catanzaro, B., 2018. High-resolution image synthesis and semantic manipulation with conditional GANs. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 8798–8807.

Xue, T., Chen, B., Wu, J., Wei, D., Freeman, W.T., 2019. Video enhancement with task-oriented flow. International Journal of Computer Vision 127 (8), 1106–1125.

Zhang, H., Goodfellow, I., Metaxas, D., Odena, A., 2019. Self-attention generative adversarial networks. In: International Conference on Machine Learning. PMLR, pp. 7354–7363.

Zhao, C., Dewey, B.E., Pham, D.L., Calabresi, P.A., Reich, D.S., Prince, J.L., 2020. SMORE: A Self-supervised Anti-aliasing and Super-resolution Algorithm for MRI Using Deep Learning. IEEE transactions on medical imaging.

Zhou, T., Tulsiani, S., Sun, W., Malik, J., Efros, A.A., 2016. View synthesis by appearance flow. In: European conference on computer vision. Springer, pp. 286–301.

Zhu, J., Yang, G., Lio, P., 2019. How can we make gan perform better in single medical image super-resolution? A lesion focused multi-scale approach. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). IEEE, pp. 1669–1673.

Zhu, J.-Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision, pp. 2223–2232.