Deep Atlas Network for Efficient 3D Left Ventricle Segmentation on Echocardiography

Suyu Dong [a], Gongning Luo [a], Clara Tam [d,e], Wei Wang [a], Kuanquan Wang [a,*], Shaodong Cao [b], Bo Chen [d,e], Henggui Zhang [a,c,*], Shuo Li [d,e,*]

[a] Biocomputing Research Center, School of Computer Science and Technology, Harbin Institute of Technology, Harbin, 150001, China
[b] The Department of Radiology, The Fourth Hospital of Harbin Medical University, Harbin 150001, China
[c] School of Physics and Astronomy, University of Manchester, Manchester, UK
[d] The Department of Medical Imaging, Western University, London, Canada
[e] The Digital Imaging Group of London, London, ON N6A 3K7, Canada

ARTICLE INFO

ABSTRACT

We proposed a novel efficient method for 3D left ventricle (LV) segmentation on echocardiography, which is important for cardiac disease diagnosis. The proposed method effectively overcame the 3D echocardiography's challenges: high dimensional data, complex anatomical environments, and limited annotation data. First, we proposed a deep atlas network, which integrated LV atlas into the deep learning framework to address the 3D LV segmentation problem on echocardiography for the first time, and improved the performance based on limited annotation data. Second, we proposed a novel information consistency constraint to enhance the model's performance from different levels simultaneously, and finally achieved effective optimization for 3D LV segmentation on complex anatomical environments. Finally, the proposed method was optimized in an end-to-end back propagation manner and it achieved high inference efficiency even with high dimensional data, which satisfies the efficiency requirement of clinical practice. The experiments proved that the proposed method achieved better segmentation results and a higher inference speed compared with state-of-the-art methods. The mean surface distance, mean hausdorff surface distance, and mean dice index were 1.52 mm, 5.6 mm and 0.97 respectively. What's more, the method is efficient and its inference time is 0.02s. The experimental results proved that the proposed method has a potential clinical application for 3D LV segmentation on echocardiography.

1. Introduction

3D left ventricle (LV) segmentation on 3D echocardiography (3DE), which directly uses the full LV volume as input, is very important for cardiac disease diagnosis. This is because echocardiography, now, is the most widely used imaging modality for assessing the cardiac structure and function (Pedrosa et al., 2016; Smistad and Lindseth, 2014; Ge et al., 2019b; 2019a), and 3D LV segmentation on 3DE has inherent advantages from its 3D spatial context information capable of providing more anatomical structure information for cardiac disease diagnosis to clinicians (Chacko and Ramlingam Rajkumar, 2017). As shown in Fig. 1, 3D LV segmentation fully exploits the spatial context information contained in the LV volume, while 2D LV segmentation only makes use of par-

tial spatial information contained in specific cross-sectional planes (Corsi et al., 2005). Hence, 3D LV segmentation on 3DE, compared with the traditional 2D LV segmentation on 2D echocardiography (2DE), is a more valuable task and has attracted much attention.

However, 3D LV segmentation on 3DE is still an open and challenging task due to the following intrinsic limitations: high dimensional data, complex anatomical environments, low image quality and limited annotation data. Specifically, (1) 3DE has higher dimensional data than 2DE, which makes achieving a very high inference efficiency difficult. (2) 3DE has complex anatomical environments (Lang et al., 2012), which makes obtaining highly expressive information for accurate segmentation problematic. (3) 3DE data have relatively lower signal-to-noise ratio compared to other imaging modalities and have motion artifacts. Furthermore, up to now, annotated 3DE data is limited, which makes it difficult to train a model to achieve high accuracy on the 3D LV segmentation task based on limited 3DE datasets. Hence, how to utilize 3DE's advan-
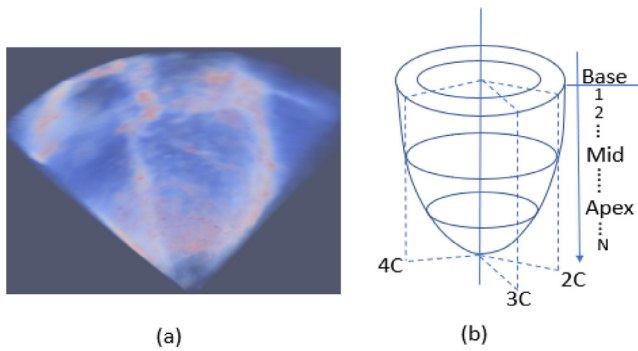
Fig. 1. 3D left ventricle volume contains more spatial context information than 2D left ventricle's typical (specific) cross-sectional planes. (a) 3D left ventricle volume on echocardiography. (b) 2D left ventricle's typical (specific) cross-sectional planes on echocardiography, which includes the planes from base to apex in short axis and 4C, 3C, and 2C planes in long axis.

tages and overcome its difficulties to get more accurate 3D LV segmentation results is an urgent problem.

Although some methods have been proposed for 3D LV segmentation on 3DE (Leung and Bosch, 2010), these methods are still difficult to overcome the above challenges completely. These methods can be classified as: traditional methods (deformable models, statistical models, graph-cuts methods, atlas-based methods), machine learning methods, and deep learning-based methods.

(1) Traditional methods: Deformable models and statistical models are widely used for 3D LV segmentation on 3DE (Santiago et al., 2015) (Pedrosa et al., 2017). Deformable models are defined by energy functions and are optimized iteratively to achieve the energy minimization for accurate delineation of the LV boundary. These models can be divided into a snake model and level-set model, which are based on the boundary and region information respectively to achieve model optimization (Yu et al., 2014). Usually, some new constraints are integrated into the common energy function to achieve a relatively more accurate segmentation of 3D LV (Zhu et al., 2010)(Zagrodsky et al., 2005). Statistical models are based on statistical information, which are mainly contour borders and image texture information, and are modeled using some specific parameters (Zhang et al., 2012). The following two statistical models: active appearance model (Mitchell et al., 2002) and active shape model (Van Assen et al., 2006) have also been widely applied for 3D LV segmentation on echocardiography and have achieved meaningful LV segmentation results (Hansegard et al., 2007; Ma et al., 2010; Barbosa et al., 2013; 2014; Orderud and Rabben, 2008). Graph-cuts were also used for 3D LV segmentation (Juang et al., 2011; Bernier et al., 2017; Mukherjee et al., 2013), which adopted graphs and graph-flow optimization techniques to solve segmentation problems (Bernard et al., 2016). Atlas-based methods, which achieve segmentation through registration techniques, are popular methods for 3D LV segmentation on 3DE (Zhuang et al., 2010b; Krishnaswamy et al., 2018; Schneider et al., 2012; Oktay et al., 2014). However, these traditional methods (deformable models, statistical models, graph-cuts and atlas-based methods) are all based on iterative on-line optimization.

(2) Machine learning methods, such as random forest and dictionary learning, have been used to segment 3D LV on echocardiography (Yang et al., 2011; Milletari et al., 2014; Lempitsky et al., 2009; Coppini et al., 1995; Huang et al., 2014). However, these methods are usually based on handcrafted features (Zhen et al., 2015). Yet, these handcrafted features have limited representation capabilities to represent the 3D LV on 3DE, which has more complex anatomical environments than 2DE.

(3) Deep learning-based methods have been successfully used for medical image analysis, especially for cardiac images(Oktay et al., 2018; Carneiro and Nascimento, 2013; Carneiro et al., 2012; Carneiro and Nascimento, 2010; Dong et al., 2016; Mo et al., 2018; Xue et al., 2017a; 2017b; 2017c; Luo et al., 2018; 2020). Besides, deep convolutional neural networks have been implemented for LV segmentation on 3D echocardiography (Oktay et al., 2018; Dong et al., 2016). These methods can automatically extract features, which have higher representation ability than handcrafted features. However, for getting features with higher representation ability and achieving better segmentation they need large numbers of annotated 3DE data, which are currently inaccessible, to learn lots of parameters. Additionally, it is worth mentioning that more and more methods try to use the deep learning technology to address the medical image registration problem (Krebs et al., 2019; Miao et al., 2016; Fan et al., 2019; Cao et al., 2018; Hu et al., 2018; de Vos et al., 2019). Besides, the atlas is considered as an important prior to improve the performance of the medical image registration task, such as the brain image registration task (Balakrishnan et al., 2019).

In this paper, inspired by the deep learning based registration methods, we proposed a novel efficient method for 3D LV segmentation on 3DE. The proposed method combined deep atlas network with information consistency constraint which overcame the above challenges effectively. The proposed deep atlas network achieved segmentation through a transformer net and deformable layers. Additionally, it was optimized in an end-to-end back propagation manner based on the proposed information consistency constraint.

The proposed method has three contributions: (1) For the first time, the proposed method explicitly integrated significant LV atlas prior into a deep learning framework to address the 3D LV segmentation problem on echocardiography. Specifically, the integrated atlas prior knowledge can accurately and completely represent the basic structure of the 3D LV. Hence, based on such basic structure, the proposed method achieves 3D LV segmentation through learning specific transformation parameters for atlas deformation. Moreover, these transformation parameters are represented as a small number of nodes, which can be trained based on limited annotation data. (2) For the first time, it used a novel multi-level information consistency constraint to enhance the model's performance from different levels simultaneously for 3D LV segmentation in complex anatomy environments. (3) What's more, the proposed couple Generator Adversarial Networks (Couple-GAN) is combined into the proposed framework to model the global semantic information.

Compared with the preliminary work in MICCAI 2018 (Dong et al., 2018), the current work has done some improvements: (1) We proposed a new couple adversarial consistency constraint (i.e., Couple-GAN). (2) We improved the volume consistency constraint through adopting a phase-based similarity measure item to further enhance the proposed method's performance. (3) We carried out more extensive experiments on performance analysis and comparison, to evaluate the proposed method and prove its application potential.

The rest of the paper is organized as follows: section 2 introduces the formulation and property of the proposed framework; the implementation details are given in section 3; we show the experiments and results in section 4; and we conclude the methods and results in section 5.
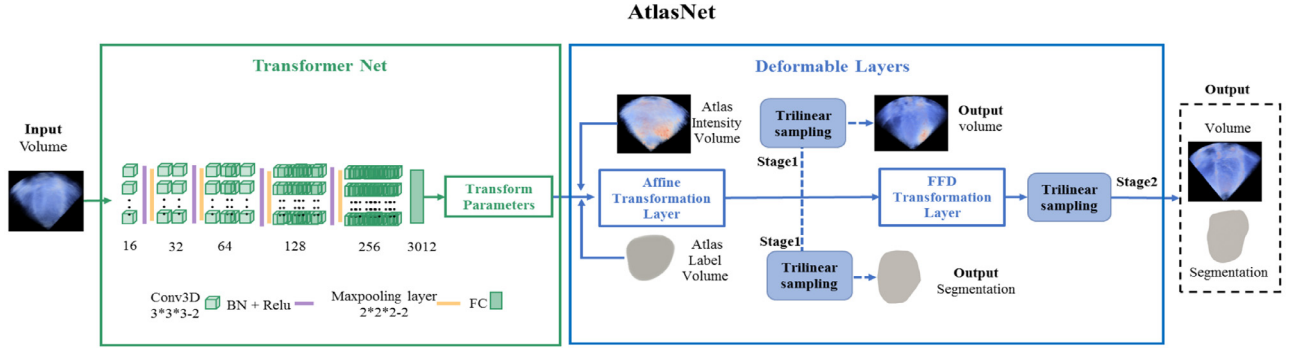
Fig. 2. The overview of the deep atlas network. The input volume is passed through the transformer net to obtain the translation parameters, which are used to warp the atlas into the deformable layers. The deformable layers then deform the atlas intensity volume and atlas label volume to the output volume and segmentation result.

2. Method

The proposed method combined deep atlas network with information consistency constraint for 3D LV segmentation on 3DE. Deep atlas network (AtlasNet), which is demonstrated in Fig. 2, advantageously integrates a powerful atlas prior into an end-to-end optimization framework (The gradient required for optimization or backpropagation can be computed). And hence it was optimized in an end-to-end back propagation manner with information consistency constraint. The details of the proposed method are described in the following sections.

2.1. Deep atlas network for spatial deformation

AtlasNet consists of two main components: transformer net and deformable layers. The transformer net was used to regress transformer parameters, and the deformable layers took advantage of transformer parameters to deform the atlas onto the target object. **Additionally, in order to avoid confusion, we denote the input volume as IV, the label of the input volume as LA, atlas intensity volume as AIV, atlas label volume as ALV, output volume as OV, and output segmentation result as OS.**

2.1.1. Transformer net for parameters regression

Transformer net learns the transformation relationship based on a large variability between IV and AIV to regress transformer parameters. The regressed transformer parameters are used to make further atlas deformation. This strategy is inspired by traditional atlas segmentation methods, which performs segmentation using a registration strategy based on the transformation relationship between the atlas and new samples. Different from the traditional atlas segmentation method, which is based on on-line optimization, the proposed transformer net enables off-line optimization with high-efficiency inference and utilizes the superiority of deep learning technology at the same time. Specially, the transformer net learns to regress the required transformer parameters by the guidance of the proposed loss function (details will be introduced in section 2.2), which measures the diversity between the transformation results and the ground truth. The structure of the transformer net is shown in Fig. 2. It has five 3D convolution layers to capture 3D context information and one fully connected layer for parameters regression. In addition, every convolution layer, except for the last convolution layer, is followed by a BN layer (Ioffe and Szegedy, 2015), Relu activation layer (Rectified linear units f(x)=max(0,x)) and max pooling layer. The receptive field sizes of every convolution layer and max pooling layer are 3*3*3 voxels and 2*2*2 voxels, and the stride is 2. The fully connected layer has 3012 nodes which mean that the number of regressed transformer parameters is 3012 (Specifically, the first 12 parame-
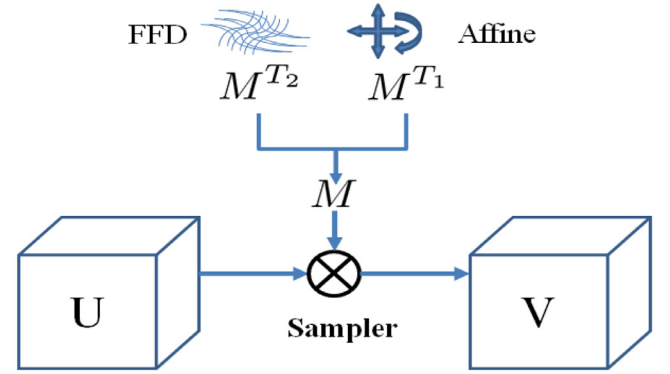

Fig. 3. A schematic diagram for the deformable layers.

ters are interpreted as the global affine transformation parameters, which are used for translation, shear, scale and rotation; the following 3000 parameters are interpreted as the moving vectors of the 1000 three-dimensional control points for free-form deformation (FFD), which are used to model the local deformation of an atlas. Hence, the meaning of these 3012 parameters are known and interpretable).

2.1.2. Deformable layers for atlas deformation

Deformable layers take advantage of the transformer parameters and warping strategy to deform the atlas (AIV and ALV) into the output (OV and OS). In this paper, the transformer parameters include $\{T_{global}, T_{local}\}$, which contain the global affine transformation parameters T_{global} and local non-rigid transformation parameters T_{local}, corresponding to affine transformation in stage 1 and free-form deformation (FFD) in stage 2 respectively. Global affine transformation achieves large-scale space registration. The local non-rigid FFD transformation achieves small-scale refinement on details to get more accurate segmentation result. Hence, as shown in Fig. 3, we modeled the deformable layers as follows:

$$
\begin{aligned}
&F = Tri\_sample(M, U), M \in \left(M^{T_1}, M^{T_2}\right), \\
&M^{T_1} = T_{global}(V), M^{T_2} = T_{local}\left(M^{T_1}\right), \\
&M_i = \left(x_i^t, y_i^t, z_i^t\right), \\
&M_i^{T_1} = \left(x_i^{t1}, y_i^{t1}, z_i^{t1}\right), M_i^{T_2} = \left(x_i^{t2}, y_i^{t2}, z_i^{t2}\right), \\
&V \in \{OV, OS\}, V_i = (x_i, y_i, z_i), \\
&U \in \{ALV, AIV\}
\end{aligned}
\tag{1}
$$

where

- F denotes the final transformation function between the atlas and output using trilinear interpolation,

- i denotes the index of a voxel from final output volume and it is consistent among OV, OS, and LA,
- M denotes the mapping matrix from the output to the atlas (Type: matrix; Dimension: $300 \cdot 300 \cdot 300$; Element: 3D vector),
- M^{T_1} denotes the mapping matrix only using the global affine transformation after stage 1 (Type: matrix; Dimension: $300 \cdot 300 \cdot 300$; Element: 3D vector),
- M^{T_2} denotes the mapping matrix using the global affine transformation and local FFD after stage 2 (Type: matrix; Dimension: $300 \cdot 300 \cdot 300$; Element: 3D vector),
- M_i denotes voxel i's mapping coordinates (x_i^t, y_i^t, z_i^t) from output V to atlas U (Type: vector; Dimension: $3 \cdot 1$; Element: scalar). In this equation, x_i, y_i, and z_i are floats.
- $M_i^{T_1}$ and $M_i^{T_2}$ denote voxel i's mapping coordinates $(x_i^{t1}, y_i^{t1}, z_i^{t1})$ and $(x_i^{t2}, y_i^{t2}, z_i^{t2})$ from stage 1 and stage 2 respectively (Type: vector; Dimension: $3 \cdot 1$; Element: scalar). In this equation, x_i^{t1}, y_i^{t1}, z_i^{t1}, x_i^{t2}, y_i^{t2}, and z_i^{t2} are floats.
- V includes the OV and the OS (Type: matrix; Dimension: $300 \cdot 300 \cdot 300$; Element: scalar),
- $V_i = (x_i, y_i, z_i)$ denotes the coordinate of voxel (or grid) i in V (Type: vector; Dimension: $3 \cdot 1$; Element: scalar). In this equation, x_i, y_i, and z_i are integers.
- U denotes the 3D LV atlas including ALV and AIV (Type: matrix; Dimension: $300 \cdot 300 \cdot 300$; Element: scalar).

The 3D LV atlas was used to provide basic anatomic structure knowledge for the proposed method. The benefit of integrating atlas into our proposed deep learning network is that the atlas is a comprehensive prior. This comprehensive prior not only reduces the data dependence but also directly provides a good prior for deep learning. This is especially important for addressing the challenge of limited data. Additionally, the comprehensive prior information is useful for addressing the challenge of indistinguishable boundaries between the LV and other cardiac substructures. Those challenges make using simple prior information for accurate segmentation problematic.

The employed warping strategy comes from the widely used spatial transformer strategy in (Jaderberg et al., 2015). This strategy warps the atlas (AIV and ALV) into output (OV and OS). As shown in Fig. 2, the following affine transformation layer (T_{global}) and FFD transformation layer (T_{local}) generate transformations between any grid V_i in V and the corresponding mapping coordinates M_i^T in atlas U (AIV and ALV). Based on the such transformation, the trilinear interpolation sampling was used to produce the warped atlas, which was used as the final 3D output. The details are shown as follows:

Trilinear interpolation sampling can be described as

$$
F_i = \sum_{a=1}^{300} \sum_{b=1}^{300} \sum_{c=1}^{300} U_{a,b,c} max\left(0, 1 - \left|x_i^t - a\right|\right)
$$
$$
\cdot max\left(0, 1 - \left|y_i^t - b\right|\right) \cdot max\left(0, 1 - \left|z_i^t - c\right|\right) \tag{2}
$$

where $U_{a,b,c}$ denotes the value in position (a, b, c) on the atlas (a, b, c are integers), F_i is the sampling value by interpolation on the atlas for voxel V_i of the output, and the i is the index of a voxel.

Affine transformation describes the global transformation from the output to the atlas. It adopts twelve degrees of freedom (including translation, shift, scale and rotation) (Rueckert et al., 1999) to achieve the basic anatomic structure mapping between the output and the atlas. The 3D affine transformation can be modeled as

$$
T_{global}(V_i) = \begin{bmatrix} \Theta_1 & \Theta_2 & \Theta_3 \\ \Theta_4 & \Theta_5 & \Theta_6 \\ \Theta_7 & \Theta_8 & \Theta_9 \end{bmatrix} \cdot V_i + \begin{bmatrix} \Theta_{10} \\ \Theta_{11} \\ \Theta_{12} \end{bmatrix} \tag{3}
$$

$\Theta_1 \cdots \Theta_{12}$ denotes the affine transformation parameters which are regressed by the transformer net, and $V_i = (x_i, y_i, z_i)$ denotes the coordinate of voxel (or grid) i in V (Type: vector; Dimension: $3 \cdot 1$).

Free-form deformation can refine any local non-rigid transformation to achieve high pixel-level accuracy to process the local variation of the LV from different people. In this paper, FFD is based on B-splines transformation(Rueckert et al., 1999), and it shifts control points to capture the deformation of the object. The number of control points defines the degree of non-rigid deformation (we used 1000 control points and each control point is three-dimensional, hence 3000 parameters were used for FFD). 3D FFD can be modeled as:

$$
T_{local}\left(M_i^{T_1}\right) = \sum_{l=0}^{3} \sum_{m=0}^{3} \sum_{n=0}^{3} B_l(u)B_m(r)B_n(w)\phi_{a+l,b+m,c+n}
$$
$$
a = \left\lfloor x_i^{t1}/\delta \right\rfloor - 1, b = \left\lfloor y_i^{t1}/\delta \right\rfloor - 1
$$
$$
c = \left\lfloor z_i^{t1}/\delta \right\rfloor - 1, u = x_i^{t1}/\delta - \left\lfloor x_i^{t1}/\delta \right\rfloor
$$
$$
r = y_i^{t1}/\delta - \left\lfloor y_i^{t1}/\delta \right\rfloor, w = z_i^{t1}/\delta - \left\lfloor z_i^{t1}/\delta \right\rfloor \tag{4}
$$

where ϕ represents the 3D control points $\phi_{a,b,c}$ with uniform spacing $\delta = 30$. The B_l denotes the lth basis function of B-splines (we adopted the specific basis functions used in (Rueckert et al., 1999)). FFD achieves local non-rigid transformation by shifting control points location from $\phi_{a+l,b+m,c+n}$ to $\phi_{a+l,b+m,c+n} + \Delta\phi_{a+l,b+m,c+n}$ which can be formulated as:

$$
T_{local}\left(M_i^{T_1}\right) = \sum_{l=0}^{3} \sum_{m=0}^{3} \sum_{n=0}^{3} B_l(u)B_m(r)B_n(w)
$$
$$
\cdot \left(\phi_{a+l,b+m,c+n} + \Delta\phi_{a+l,b+m,c+n}\right) \tag{5}
$$

In this paper, to achieve non-rigid deformation, the regressed transformer parameters from transformer net for FFD are $\Delta\phi_{a+l,b+m,c+n}$ of every control point.

2.2. Optimization object with multi-level information consistency constraint

As shown in Fig. 4, the multi-level information consistency constraint was proposed to advantageously capture multi-level representative information to improve the segmentation accuracy. It includes label consistency constraint, volume consistency constraint and adversarial consistency constraint. Label consistency constraint can model the loss from the local semantic information of the anatomical structure of the 3D LV based on the consistency between the output segmentation result and the 3D LV label. Volume consistency constraint enhances the learning of shared latent transformer variables based on the consistency between the OV and IV. During model optimization, when the segmentation consistency constraint cannot converge, the volume consistency constraint can provide constraint to make the model converge. Label consistency constraint and volume consistency constraint both focus on voxel-level loss modeling. Adversarial consistency constraint tries to make the results look realistic based on global-level semantic information. Therefore, multi-level information consistency constraints are of great benefit to 3D LV segmentation on complex anatomical environments. The final optimization object of the proposed framework is defined by the following formula:

$$
L = \alpha L_{label} + \beta L_{intensity} + L_{image-cGAN}(G, D) \tag{6}
$$

where L_{label}, $L_{intensity}$, and $L_{image-cGAN}$ are label consistency constraint, volume consistency constraint, and adversarial consistency constraint respectively. The weights of L_{label} and $L_{intensity}$ are α and β. The details of the three-fold information consistency constraint are described in the following.
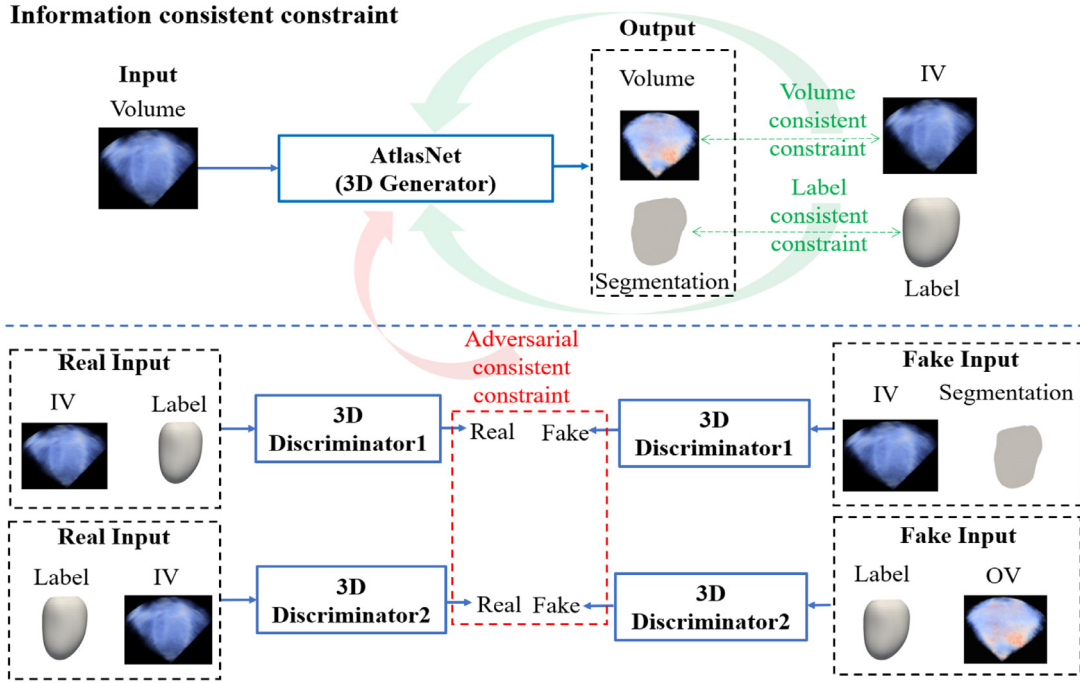
Information consistent constraint

Fig. 4. Multi-level information consistency constraint is used for optimization based on the image-conditional GAN. Label consistency constraint and volume consistency constraint can model voxel-level loss with local semantic information. Adversarial consistency constraint can model global-level loss based on global-level semantic information.

2.2.1. Label consistency constraint

Label consistency constraint is an important part of the loss function. It is used to measure the similarity between the 3D LV label and the OS to guide the atlas deformation during model optimization. It was modeled by ridge regression:

$$L_{label} = \frac{1}{300 \cdot 300 \cdot 300} \sum_{x=1}^{300} \sum_{y=1}^{300} \sum_{z=1}^{300} \left\| LA_{(x,y,z)} - OS_{(x,y,z)} \right\|_2 \qquad (7)$$

where (x, y, z) denotes the coordinate of a voxel in LA and OS (In this equation, the voxel coordinates of the LA and OS are integers). Although label consistency constraint encourages the OS to correspond with the 3D LV label, the OS is blurry when only using the label consistency constraint. Because the nature of ridge regression is to predict the mean of the distribution, which will reduce the overall loss but output blurry results (Pathak et al., 2016). In this paper, the blurry results were refined by the following adversarial consistency constraint.

2.2.2. Volume consistency constraint

Volume consistency constraint is another important part of the loss function. It is used to measure the similarity between the OV and IV. It is designed based on the fact that the deformable layers are shared between the OS and OV. Because of the consistency of optimization variables for deformable layers, adding volume consistency constraint can enhance the learning of shared latent transformer variables. So, the volume consistency constraint was proposed as part of the loss function to guide the atlas deformation. Due to the large random noise and high intensity variability on 3DE volumes, the volume consistency constraint was modeled through a phase-based similarity measure method (Zhang et al., 2007), which has high robustness for the intensity variations from ultrasound images (Zhang et al., 2007). Following the phase-based similarity measure method in (Zhang et al., 2007), the volume con-

sistency constraint is formulated as:

$$L_{intensity} = -\sum_{j=1}^{J} \sum_{k=1}^{K} \mathbb{P}(\phi_j(IV), \phi_k(OV)) \cdot ln\left(\frac{\mathbb{P}(\phi_j(IV), \phi_k(OV))}{\mathbb{P}(\phi_j(IV)) \cdot \mathbb{P}(\phi_k(OV))} \right)$$

$$(8)$$

where J and K denote the number of bins of the phase probability histogram from IV and OV respectively, j and k denote the index of a bins, ϕ is the phase computation function, $\phi_j(IV)$ denotes the phase of the IV in j, $\phi_k(OV)$ denotes the phase of the OV in k, $\mathbb{P}$ is the phase probability density function, $\mathbb{P}(\phi_j(IV), \phi_k(OV))$ is the joint phase probability of $\phi_j(IV)$ and $\phi_k(OV)$, $\mathbb{P}(\phi_j(IV))$ is the individual phase probability of $\phi_j(IV)$, and $\mathbb{P}(\phi_k(OV))$ is the individual phase probability of $\phi_k(OV)$. Specifically, the phase function ϕ is defined based on the widely used phase computation method in (Mellor and Brady, 2005) (which uses the scale invariant filter to estimate phase, and also shows that the phase computation step is a continuously differentiable function).

2.2.3. Adversarial consistency constraint

Adversarial consistency constraint is based on the image-conditional Generative Adversarial Nets (GAN) which is appropriate for image-to-image translation tasks (Isola et al., 2017). Image-conditional GAN is a conditional GAN which uses an image (or volume in 3D) as input condition. Conditional GAN is extended from GAN and learns a conditional generative model (Mirza and Osindero, 2014). In conditional GAN, its generator G learns the mapping from condition p and random noise vector q to r', $G : \{p, q\} \to r'$ and tries to make the output look real through minimizing $E_{p,r'}\left[\log\left(1 - D\left(p, r'\right)\right)\right]$; its discriminator D tries to distinguish the generated r' and the ground truth r through maxing the probability of the correct label of r and r' (Goodfellow et al., 2014). D outputs a single scalar, which represents the probability of the discriminator's input that came from the r rather than r' (Goodfellow et al., 2014). In the training process, the discriminator and generator are trained iteratively. When we train the discriminator, the genera-

tor's parameters are fixed. When we train the generator, the discriminator's parameters are fixed. Besides, the discriminator provides loss gradients to the generator (Goodfellow et al., 2014). It is a two-player game and its objective is:

$$L_{cGAN}(G, D) = E_{p,r}[logD(p, r)] + E_{p,r'}\left[log\left(1 - D\left(p, r'\right)\right)\right] \quad (9)$$

where $E_{p,r}$ denotes the expectation of the function $logD(p, r)$ based on the output probability of the discriminator about the real subject r with condition p, and $E_{p,r'}$ denotes the expectation of the function based on the output probability of the discriminator about the generated subject r' with condition p. Analogous to conditional GAN, the image-conditional GAN's generator tries to produce a 'real' image (or volume) to confuse the discriminator, and its discriminator tries to detect the 'fake' from the generator. Image-conditional GAN used in this paper is similar to the research in (Isola et al., 2017). The discriminator pays attention to the structural information in an image. Hence, image-conditional GAN automatically learns a structured loss function, which models global-level semantic information to penalize any possible structural differences between the output and target (Isola et al., 2017). Additionally, different from the label consistency constraint and volume consistency constraint, which focuses on voxel-to-voxel consistency constraints, the learned structured loss from GAN can penalize the joint configuration of the output (Isola et al., 2017). This means that adversarial consistency constraint takes advantage of the global-level semantic information to further refine the segmentation results. Hence, adversarial consistency constraint was employed by using AtlasNet as the generator G of image-conditional GAN framework. Prior work in (Dong et al., 2018) used one generator and one discriminator to model adversarial consistency constraint, and the discriminator was used for detecting the fake and real label. However, different from prior work in (Dong et al., 2018), we proposed a Couple-GAN (as shown in Fig. 4) with one generator and a couple of discriminators to improve the adversarial consistency constraint on the OS and OV respectively. In this way, the adversarial consistency constraint discriminated the OS as well as the OV simultaneously. Hence, the adversarial consistency constraint was further improved to boost the performance of the proposed method. The adversarial consistency constraint is expressed as:

$$L_{image-cGAN}(G, D_1)$$
$$= E_{IV,LA}[logD_1(IV, LA)] + E_{IV,OS}[log(1 - D_1(IV, OS))] \quad (10)$$

$$L_{image-cGAN}(G, D_2)$$
$$= E_{LA,IV}[logD_2(LA, IV)] + E_{LA,OV}[log(1 - D_2(LA, OV))] \quad (11)$$

$$G(IV) = [OV, OS] \quad (12)$$

$$L_{image-cGAN}(G, D) = L_{image-cGAN}(G, D_1) + L_{image-cGAN}(G, D_2) \quad (13)$$

Both generator (AtlasNet) G and discriminators D_1, D_2 are end-to-end jointly optimized using stochastic gradient descent.

2.3. End-to-end optimization

In the proposed novel framework, the gradient required for optimization from the deformable layer and the information consistency constraint loss function can be computed. As a result, the loss gradient can flow back to optimize the parameters of the proposed framework. Hence, the proposed framework was optimized in an end-to-end manner. Specifically, the gradient compuation method from trilinear sampling and affine transformation has been proven in (Jaderberg et al., 2015), and the gradient compuation method from free-form deformation has been proven in

Table 1
The structure of the discriminator.

Symbol	Kernel (kernel size-stride)	Number of channels
Conv3D	7*7*7-2	16
BN + Relu	-	-
Conv3D	5*5*5-2	32
BN +Relu	-	-
Conv3D	3*3*3-2	64
BN +Relu	-	-
Conv3D	3*3*3-2	128
BN +Relu	-	-
Conv3D	3*3*3-2	256
Sigmoid	-	-

(Zhang and He, 2017). The gradient compuation method from the label consistency constraint, volume consistency constraint, and adversarial consistency constraint has been proven in (Wu et al., 1996), (Mellor and Brady, 2005), and (Goodfellow et al., 2014) respectively.

2.4. Network architecture

Due to the adversarial consistency constraint employed in this paper, the network architecture of the proposed segmentation method is an image-conditional GAN framework, which has been illustrated in Fig. 4. The proposed AtlasNet was used as the generator G and its detailed structure has been shown in Fig. 2. Additionally, the transformer net of the AtlasNet is flexible and can also use other structures, such as 3D Resnet 101 (Ioffe and Szegedy, 2015) and 3D DenseNet 121 (Huang et al., 2017), which require more convolution computation. 3D Resnet and 3D DenseNet structures have been experimented as the transformer net for 3D LV segmentation on the proposed framework. However, they did not achieve a better performance, used more than 10^2 times of Flops and was more susceptible to over-fitting than the proposed structure in Fig. 2. Hence, the relatively complex and heavy-weight network structure is not proper for the proposed framework. In order to make a proper compromise between the inference efficiency and the performance, the transformer net used the proposed structure in Fig. 2 in this paper.

The discriminators D_1 and D_2 have the same structure and have been shown in Table 1. But, their inputs are different. D_1 used the IV and corresponding label as the real input, and used the IV and corresponding OS as the fake input. D_2 used the corresponding label and IV as the real input, and used the corresponding label and OV as the fake input. These two discriminators all output the judgment for the input.

3. Implementation Details

3.1. Atlas

In the proposed framework, only a single atlas was used to provide basic anatomical structure knowledge. Compared with multi-atlas, the advantage of the single atlas is that it does not need to select atlases and fuse labels, which are essential for the multi-atlas based method to obtain good segmentation accuracy. Hence, a single atlas is more efficient than the multi-atlas. For the left ventricle segmentation on 3DE, efficiency is also very important. Hence, we used the single atlas in the proposed framework. The cost is that atlas (which includes an intensity volume and a corresponding shape label) need to be constructed in an independent process. To ensure a fair comparison among other methods, the single atlas was built from the ten 3DE volumes and the corresponding ten labels, which were randomly selected from the training set. The selected volumes and labels were firstly registered to

a mean space. Then, the atlas intensity volume and atlas label volume were computed based on the registered 3DE volumes and the corresponding labels respectively. The details of the atlas construction can be referred to in (Zhuang et al., 2010a). In the inference stage, based on the atlas and the regressed translation parameters, we can achieve the final 3D LV segmentation.

Comparison with the state-of-the-art atlas-based segmentation method on 3DE is important to evaluate the proposed method. In this paper, we compared the proposed method with the multi-atlas segmentation method in (Oktay et al., 2014), which is the best atlas-based method for LV segmentation on 3DE. In the multi-atlas construction procedure and segmentation procedure of (Oktay et al., 2014), atlas selection was achieved through selecting the most similar 10 atlases ((Oktay et al., 2014) proved that 10 atlases can achieve the best segmentation performance) from the training set based on a normalized mutual information similarity metric, and label fusion was achieved by majority voting of the labels generated from the 10 atlases.

3.2. Training process

The training process of the proposed method has been summarized in Algorithm 1. Due to the proposed method being an image-conditional GAN method, we must train the framework using an iterative approach(Goodfellow et al., 2014). We initialized the weights of the generator and discriminators using the same initialization method in (He et al., 2015).

Algorithm 1 The training process of the proposed method

---

1. Initializing the network parameters.
2. **While** $mu < MIN_{step1}$:
3.   Training AtlasNet with the affine transformer parameters $\Theta_1....\Theta_{12}$ using
        Back Propagation (BP), without adversarial consistency constraint, and fixing parameters $\Theta_{13}....\Theta_{3012}$.
        $mu = mu + 1$.
4. **End while**
5. $mu = 0$.
6. **While** $mu < MIN_{step2}$:
7.   Training AtlasNet with the transformer parameters $\Theta_1....\Theta_{3012}$ using BP,
        without adversarial consistency constraint.
        $mu = mu + 1$.
8. **End while**
9. $mu = 0$.
10. **For** $cu < MIN_{step3}$:
11.   **While** $mu < MIN_D$:
12.   Training the discriminators using BP, to maximize
            $E_{IV,LA}[\log D_1(IV, LA)] + E_{IV,OS}[log(1 - D_1(IV, OS))]$
            $+ E_{LA,IV}[log D_2(LA, IV)] + E_{LA,OV}[log(1 - D_2(LA, OV))]$.
            $mu = mu + 1$.
13.   **End while**
14. $mu = 0$.
15.   **While** $mu < MIN_G$:
16.   Training the generator (AtlasNet) using BP, with the full three-fold information consistency constraint, to minimize
            $L = \alpha L_{label} + \beta L_{intensity} + E_{IV,OS}[log(1 - D_1(IV, OS))]$
            $+ E_{LA,OV}[log(1 - D_2(LA, OV))]$.
            $mu = mu + 1$.
17.   **End while**
18. $mu = 0$.
19. $cu = cu + 1$.
20. **End For**

The training process of 3D Generator (AtlasNet) was divided into three steps. To avoid over-fitting on the training generator (AtlasNet), in the first and second step, the affine transformer parameters and the FFD parameters were trained progressively. Specifically, in the first step, we only trained the affine transformer parameters based on the output of stage 1, fixing FFD parameters. In the second step, all the parameters were trained together based on the final output from stage 2. Besides, in the first two training steps, we only used label consistency constraint and volume consistency constraint to guide the model optimization. In the third step, the iterative training mode was used to train the generator (AtlasNet) using the full three-fold information consistency constraint.

Additionally, we trained the discriminator using the same training mode with (Isola et al., 2017), which uses a patch-driven manner to train the discriminator. In this paper, the patch-driven manner can avoid over-fitting for the discriminator. The size of patch was 70 · 70 · 70.

In each step, we adopted the stochastic gradient descent algorithm to train the model and used random rotation to achieve data augmentation during every epoch. Specifically, for every training data, the random rotations between 0 to 360 degrees along the Z axis were used to achieve data augmentation after every epoch. The learning rate and momentum were fixed. The learning rate was 0.0002, the momentum was 0.5, and the batch size was 1. The weights α and β for label consistency constraint, volume consistency constraint were 0.6 and 0.4 respectively. The MIN_{step1}, MIN_{step2}, MIN_{step3}, MIN_D, and MIN_G were 10000, 10000, 5, 10000, and 10000 respectively. The proposed method was implemented on a Pytorch framework and NVIDIA Titan X GPU on Dell T7920 (GPU is Core I7, and memory size is 24GB) for both training and application.

4. Results and discussion

4.1. Data

Our proposed method was evaluated on 3DE with 25 training subjects, 10 validation subjects and 35 testing subjects. Specifically, the validation dataset was used to avoid overfitting and obtain the optimal model. The data was acquired by Philips iE33 using X7-2T. Each subject includes two labeled volumes in the end-systole (ES) and end-diastole (ED) frames. To prove the robustness of the proposed framework for distortion, we use data from different kinds of patients (including healthy patients, myocardial infarction, myocardial hypertrophy, ventricular dilatation) who tend to have large variations or distortions among different subjects on cardiac geometry. Each volume was normalized into the unified size of 300 · 300 · 300 as input for the proposed method. The voxel size of the 3D cardiac echo acquisitions is: 0.35mm-0.85mm along x axis, 0.37mm-0.79 mm along y axis, 0.32mm-0.71 mm along z axis. The range of the voxel size after resampling to 300 is: 0.37mm-0.72mm along x axis, 0.38mm-0.81 mm along y axis, 0.33mm-0.66 mm along z axis. The data sets were acquired where the LV was completely in the field-of-view to ensure the acquired data is more standardized by forcing manual adjustment for the probe position. The ground truth was obtained by three cardiologists with more than 20 years experiences and the final labels were mutually authenticated. Specifically, to insure the ground-truth segmentation is as correct as possible, the annotation process with mutual agreement includes two parts, i.e., annotation with discussion and annotation with voting: (1) Annotation with discussion: The two face-to-face discussions were conducted to avoid variability on the annotation methods or on special cases respectively as much as possible. Specifically, after the first face-to-face discussion, the annotation method was agreed upon. The

three clinicians independently delineated each slice by considering the 3D information of its neighboring slices. To further check labeling variablity, on the second face-to-face discussion, we analyzed the labeling variability on each data. For the main variability (such as labels on the apex and base part of the ventricle), the annotation rule was discussed and agreed upon. (2) Annotation with voting: We used a voting strategy to reduce the inter-expert labeling variability and intra-expert labeling variability simultaneously. Specifically, to reduce inter-expert labeling variability, we provided quantitative information about the mutual agreement of the reference annotations from the three clinical experts in the following equation:

$$\begin{cases} LA_{x,y,z} = 1, if \sum_{Expert=1}^{3} LA_{x,y,z}^{Expert} \geq 2 \\ LA_{x,y,z} = 0, if \sum_{Expert=1}^{3} LA_{x,y,z}^{Expert} < 2 \end{cases} \quad (14)$$

where $LA_{x,y,z}$ denotes the final labeled voxel value (The left ventricle region is labeled as 1, and the background is labeled as 0) on coordinate (x, y, z), $LA_{x,y,z}^{Expert}$ denotes the labeled voxel value on coordinate (x, y, z) from one expert. To reduce the intra-expert variability, the $LA_{x,y,z}^{Expert}$ can be obtained by voting based on the 5 label results from the same expert. The detailed computation method is

$$\begin{cases} LA_{x,y,z}^{Expert} = 1, if \sum_{Time=1}^{5} LA_{x,y,z}^{Time} \geq 3 \\ LA_{x,y,z}^{Expert} = 0, if \sum_{Time=1}^{5} LA_{x,y,z}^{Time} < 3 \end{cases} \quad (15)$$

$LA_{x,y,z}^{Time}$ denotes the labeled voxel value on coordinate (x, y, z) from one time. In this way, inter-expert labeling variability and intra-expert labeling variability can be reduced simultaneously to some extent.

4.2. Evaluation metrics

It is natural to use the widely recognized criterion in (Bernard et al., 2016) to evaluate our proposed method, and the criterions include the mean surface distance (MSD), mean hausdorff surface distance (HSD), and mean dice index (D). Additionally, we also evaluated the clinical index through the correlation and Bland-Altman (BA) analysis. The differences between the reference annotations from different cardiologists are characterized according to MSD, HSD, and D. Specifically, the mean differences between all the reference annotations from cardiologist 1 and cardiologist 2 are 0.12mm (MSD), 0.14mm (HSD), and 0.993 (D) respectively. The mean differences between all the reference annotations from cardiologist 1 and cardiologist 3 are 0.11mm (MSD), 0.1mm (HSD), and 0.995 (D) respectively. The mean differences between all the reference annotations from cardiologist 2 and cardiologist 3 are 0.09mm (MSD), 0.13mm (HSD), and 0.992 (D) respectively. Hence, the annotations from different cardiologists are relatively consistent and the variability is low.

4.3. Comparison and ablation experiments

To evaluate the rationality of the proposed framework, we compared the proposed method, which combines the proposed deep atlas network with the novel information consistency constraint, with other methods (Including state-of-the-art methods and ablation models.). In order to give a better understanding of the experiments, we define the compared methods as follows:

- **3D atlas** is the method in (Oktay et al., 2014), which is a multi-atlas based method and the best atlas-based method for LV segmentation on 3DE.
- **ACNNs** is the method in (Oktay et al., 2018), which is the best current method for the 3D LV segmentation on 3DE.
- **3D CNN** is V-net (Milletari et al., 2016), which is the best open source method for 3D medical images segmentation.

Table 2
The performance comparison among the proposed method with dierent atlases. For all the methods we have performed the conventional statistical T-Test and it showed that the differences between all the results reported in this table are statistical significant (p<0.01). The Atlas5 (variance: σ) means that the Atlas5 with different levels of specklenoise (the mean is zero and the variance is σ).

Methods	D	MSD(mm)	HSD(mm)
Atlas1	0.973 ± 0.011	1.5 ± 0.29	5.53 ± 1.36
Atlas2	0.972 ± 0.009	1.51 ± 0.31	5.41 ± 1.32
Atlas3	0.974 ± 0.013	1.49 ± 0.28	5.71 ± 1.29
Atlas4	0.972 ± 0.01	1.5 ± 0.3	5.62 ± 1.38
Atlas5(used subsequently)	0.973 ± 0.012	1.52 ± 0.32	5.6 ± 1.35
Atlas5(variance:0.01)	0.969 ± 0.013	1.54 ± 0.23	5.7 ± 1.63
Atlas5(variance:0.02)	0.963 ± 0.011	1.51 ± 0.31	5.9 ± 1.52
Atlas5(variance:0.03)	0.965 ± 0.045	1.62 ± 0.54	6.1 ± 1.02
Atlas5(variance:0.04)	0.963 ± 0.035	1.64 ± 0.45	6.2 ± 1.41
Atlas5(variance:0.05)	0.96 ± 0.012	1.63 ± 0.34	6.6 ± 1.11

- **3D cGAN** denotes 3D image-conditional GAN built by V-net (which was used as the generator) and the proposed first discriminator in this paper.
- **AtlasNet** denotes the net, which is the proposed method with label consistency constraint and volume consistency constraint, but without adversarial consistency constraint. In this way, the effect of the adversarial consistency constraint can be clearly validated.
- **AtlasNet-I** denotes the net, which is the AtlasNet without volume consistency constraint. In this way, the effect of the volume consistency constraint can be clearly validated.
- **Proposed method-I (P-I)** denotes the net, which is the proposed method without volume consistency constraint. In this way, the effect of the volume consistency constraint can be further validated.
- **Proposed method-LI (P-LI)** denotes the net which is the proposed method with adversarial consistency constraint, without label consistency constraint and volume consistency constraint. In this way, the effect of the volume consistency constraint and label consistency constraint can be further validated.
- **Proposed method-D1 (P-D1)** denotes the net which is the proposed method with label consistency constraint, volume consistency constraint and the adversarial consistency constraint from the second discriminator, without equation (10). In this way, the effect of the first discriminator can be clearly validated.
- **Proposed method-D2 (P-D2)** denotes the net which is the proposed method with label consistency constraint, volume consistency constraint and the adversarial consistency constraint from the first discriminator, without equation (11). In this way, the effect of the second discriminator can be clearly validated.
- **The previous method** is the method in (Dong et al., 2018).

4.4. Analysis of the impact on atlas

Atlas is integrated into the proposed framework to provide prior knowledge. There are some differences among different atlases. In order to analyze the impact of atlases for the proposed method, we compared the performance of the proposed framework with different atlases. These atlases were constructed by randomly selecting ten different volumes and the corresponding labels from the training set for five times. As shown in Table 2, there are few performance differences among different atlases used in the proposed framework. These results suggested that the proposed method is robust for atlas's changes to some extent. The robustness comes from two aspects: 1) Atlas is used to provide basic anatomical structure knowledge of the 3D LV, which can be pro-

Table 3

The performance comparison among existing methods and the proposed method under different configurations. For all the methods we have performed the conventional statistical T-Test and it showed that the differences between all the results reported in this table are statistical significant ($p < 0.01$).

| Methods	D	MSD(mm)	HSD(mm)
3D Atlas (Oktay et al., 2014)	0.88 ± 0.03	2.26 ± 0.74	9.92 ± 2.16
ACNNs (Oktay et al., 2018)	0.94 ± 0.02	1.95 ± 0.32	6.92 ± 1.12
3D CNN (Milletari et al., 2016)	0.89 ± 0.035	2.1 ± 0.71	9.79 ± 8.9
3D cGAN	0.914 ± 0.028	1.98 ± 0.65	8.91 ± 7.3
AtlasNet-I	0.91 ± 0.027	1.99 ± 0.64	8.92 ± 7.16
AtlasNet	0.94 ± 0.021	1.81 ± 0.59	7.66 ± 5.4
P-LI	0.94 ± 0.022	1.91 ± 0.49	7.95 ± 4.37
P-I	0.95 ± 0.021	1.86 ± 0.42	7.33 ± 2.1
P-D1	0.96 ± 0.015	1.79 ± 0.36	6.51 ± 1.9
P-D2	0.96 ± 0.014	1.65 ± 0.35	6.1 ± 1.6
The previous method(Dong et al., 2018)	0.95 ± 0.019	1.85 ± 0.43	7.26 ± 2.3
Proposed method	**0.97 ± 0.012**	**1.52 ± 0.32**	**5.6 ± 1.35**
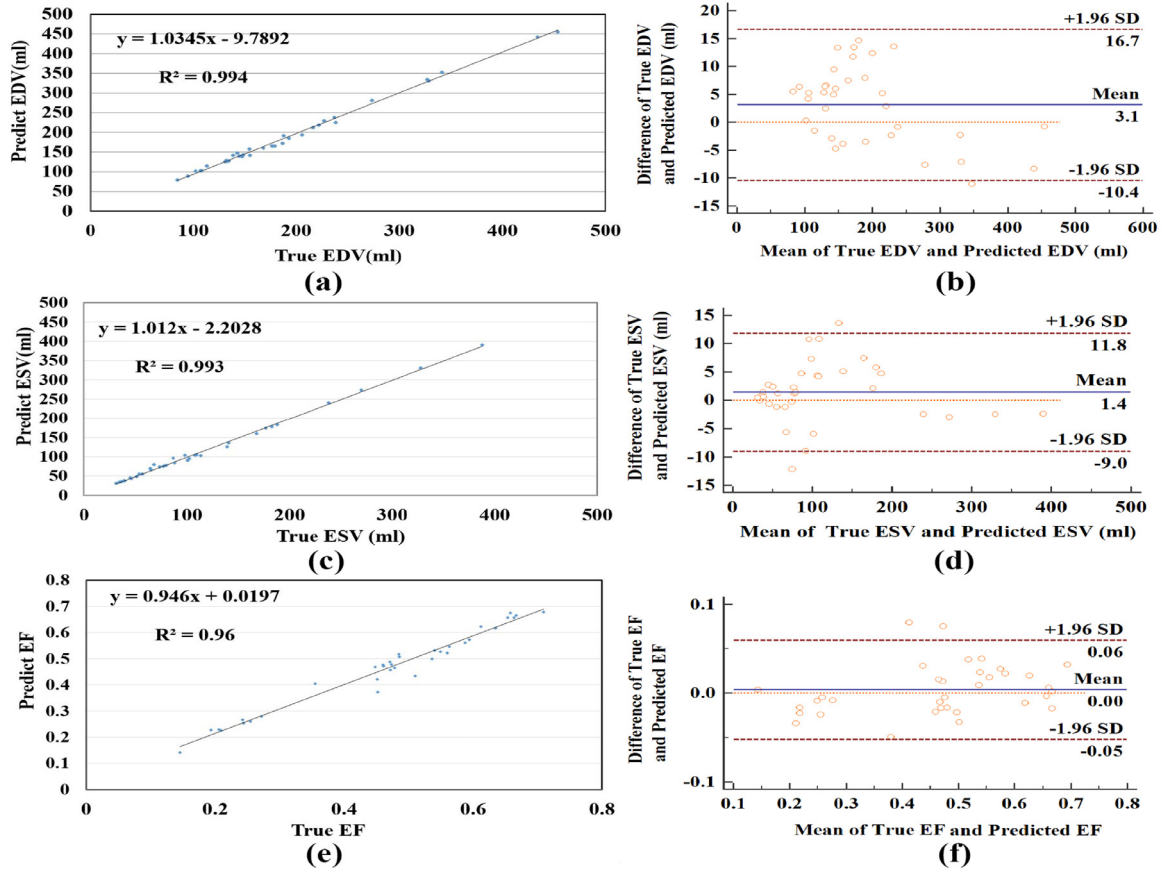
Fig. 5. Correlation graphs and Bland-Altman graphs. (a) The correlation between the predicted EDV and the true EDV. (b) The Bland-Altman analysis between the predicted EDV and the true EDV. (c) The correlation between the predicted ESV and the true ESV. (d) The Bland-Altman analysis between the predicted ESV and the true ESV. (e) The correlation between the predicted EF and the true EF. (f) The Bland-Altman analysis between the predicted EF and the true EF.

vided by any atlas; 2) The proposed method inherently is able to train the transformer net for accurate deformation to fit different atlases, hence it is robust for the atlas choice. In the following experiments, the Atlas5 was used in the proposed method.

Additionally, in order to analyze the impact of the quality of the atlas, we used the degenerated Atlas5 to conduct experiments. Specifically, we degenerated the atlas intensity volume through adding different levels of speckle noise (which is common for echocardiography) with zero mean (the variance ranges from 0.01 to 0.05). As shown in Table 2, until the variance is 0.05, the proposed method still achieved good segmentation performance (Despite the performance decreases slightly along with the increase of noise). This re-

sult demonstrated the proposed framework is relatively robust for atlas selection and quality difference.

4.5. Analysis of performance

The proposed method is able to achieve relatively better LV segmentation performance on 3DE than the state-of-the-art competitors, as shown in Table 3. The mean surface distance, hausdorff surface distance and dice were 1.52 mm, 5.6 mm and 0.97 respectively. These results prove the superiority of the proposed method compared with the existing state-of-the-art methods.
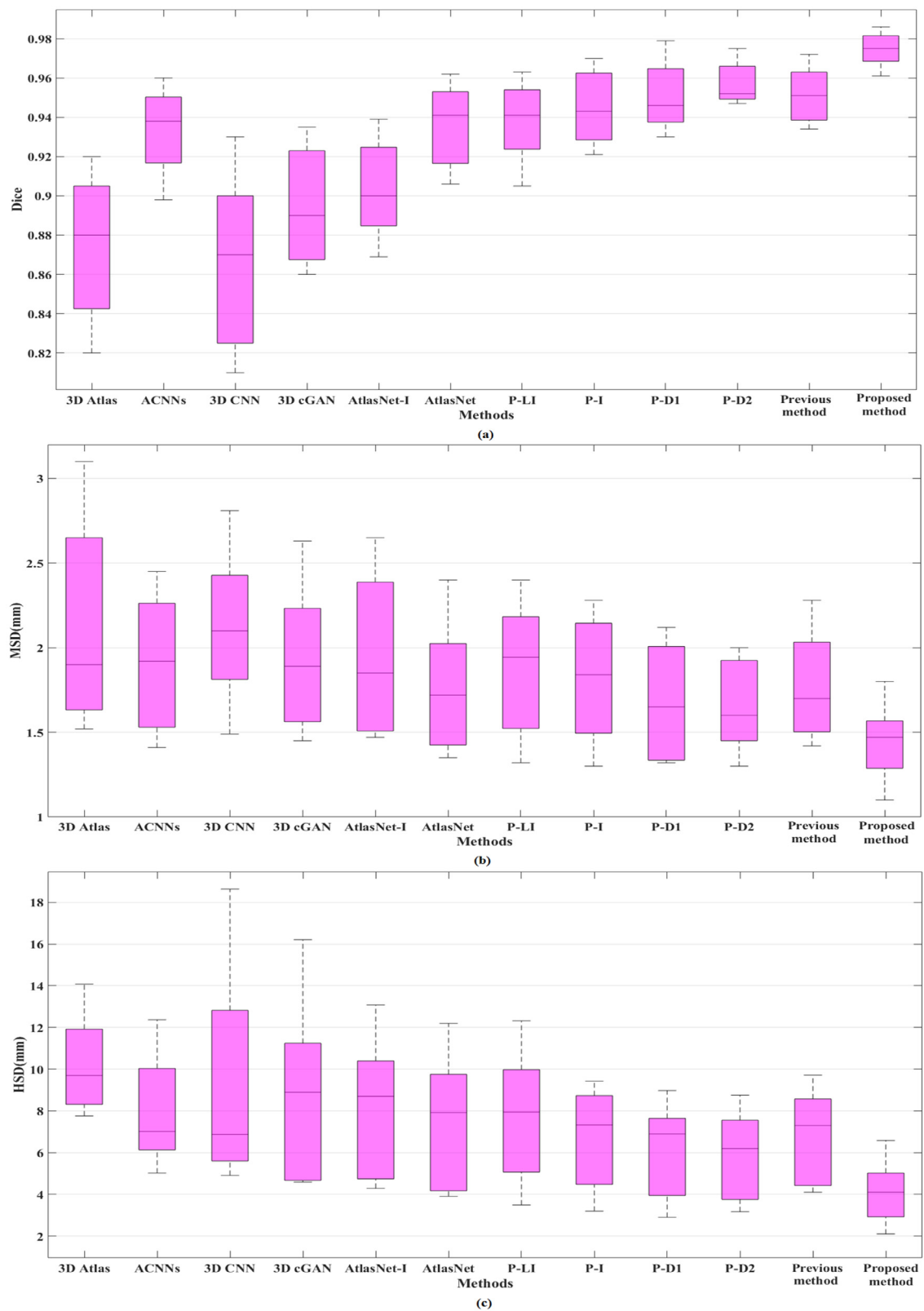
Fig. 6. The segmentation performance comparison among different methods. For all the methods we have performed the conventional statistical T-Test and it showed that the differences between all the results reported in this figure are statistical significant (p<0.01). (a) The segmentation performance Dice metric box-plot. (b) The segmentation performance MSD metric box-plot. (c) The segmentation performance HSD metric box-plot.
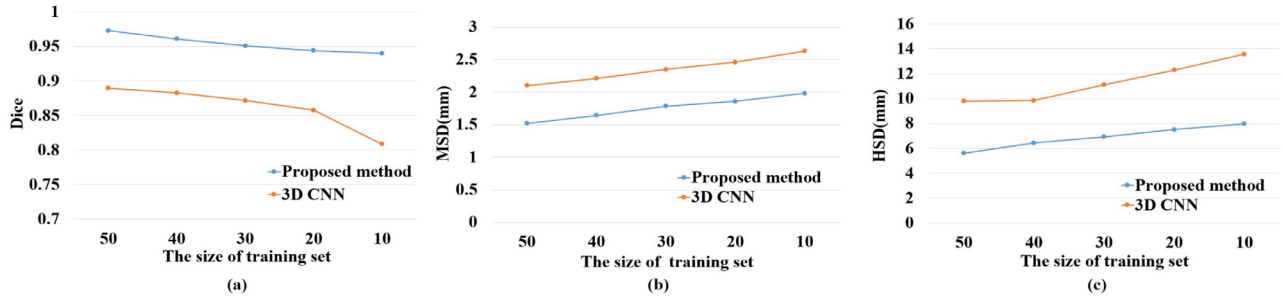
Fig. 7. The segmentation results comparison between the proposed method and 3D CNN (Milletari et al., 2016) using different sized training sets. The size of the training dataset varies from 50 to 10, and 50 corresponds to 25 subjects with 50 volumes (every subject contains an ED volume and an ES volume). For all the methods we have performed the conventional statistical T-Test and it showed that the differences between all the results reported in this figure are statistical significant (p<0.01).

Besides, compared with the existing state-of-the-art methods, the proposed method-LI and AtlasNet are still advantageous even with some ablations. This proves that combining the atlas into a CNN framework inherently has significant superiority based on the atlas' prior knowledge. Additionally, we found that the volume consistency constraint term has special value, since the AtlasNet achieved a better performance than the AtlasNet-I and the proposed method was better than the Proposed method-I. The results proved that the segmentation consistency and the volume consistency are complementary and the volume consistency constraint helped the model achieve a better performance. What's more, the proposed Couple-GAN is meaningful because the proposed method can achieve a better performance than the AtlasNet. The second discriminator was also beneficial for the proposed framework, because Proposed method-D1 performed better than AtlasNet, and the proposed method performed better than Proposed method-D2. Finally, compared with all the ablation models, the proposed method achieves the best segmentation results. This proves the reasonability of the proposed framework, which combines the full three-fold information consistency constraints.

Fig. 5 shows the clinically quantitative performance of the proposed method. The correlation graphs in Fig. 5 (a), (c), and (e) show the correlation between the ground truth and segmentation results. The correlation coefficients of EDV, ESV and EF are 0.994, 0.993, and 0.96 respectively. Additionally, the BA graphs in Fig. 5 (b), (d), and (f) show the distribution of differences between the ground truth values and segmentation values along the means between the ground truth values and segmentation values. The means of differences achieve 3.1ml (with confidence intervals between -10.4 and 16.7), 1.4ml (with confidence intervals between -9.0 and 11.8), and 0 (with confidence intervals between -0.05 and 0.06) in terms of EDV, ESV, and EF respectively.

Fig. 6 shows a performance distribution box-plot of the proposed method and state-of-the-art methods on the testing set. It proves the superiority of the proposed method compared with the state-of-the-art methods in a more representative way.

Fig. 8 and Fig. 9 show the qualitative performance of the proposed method. Fig. 8 displays the detailed comparison between the ground truth and the output label in the 2D slices from a 3D volume. Fig. 9 displays the detailed comparison between the input volume and the output volume in the 2D slices from a 3D volume. Fig. 8 proves that the proposed method was the relatively better method compared with state-of-the-art methods in the aspect of segmentation consistency and accuracy. Besides, it also shows that the atlas prior and the proposed three-fold information consistency constraint are important for model improvement. Fig. 9 shows that the output volume can be accurately registered on the input volume, which proves that the proposed method obtained accurate transformer parameters on the deformation layer to achieve the good mapping between the atlas and the input volume.

In brief, the proposed method achieved the best performance, benefitting from the proposed deep atlas network and information consistency constraint as following three points: (1) The deep atlas network integrates atlas prior knowledge, which directly comes from real data, into an end-to-end deep learning framework. Compared with the learned knowledge only obtained by a CNN, the integrated atlas is able to provide more proper prior knowledge that is more direct and accurate. This point is demonstrated by the fact that the segmentation performance using AtlasNet is better than 3D CNN in (Milletari et al., 2016). (2) The deep learning technology is more suited for transformer parameters regression for atlas deformation than the traditional atlas segmentation method. The segmentation performance using AtlasNet was better than the traditional 3D atlas segmentation method in (Oktay et al., 2014). (3) The full three-fold information consistency constraint simultaneously models the optimization object based on a multi-level loss, which is of greater benefit to 3D LV segmentation on complex anatomical environments. The L_{label} and $L_{intensity}$ constraints can model voxel-level loss and reduce the overall loss, and encourage the results with respect to the corresponding targets. $L_{image-cGAN}(G, D)$ constraint can model the loss based on global-level semantic information and further refine the segmentation results to make the results look real. Hence, the proposed method uses a multi-level information consistency constraint, combining the L_{label}, $L_{intensity}$ and $L_{image-cGAN}(G, D)$, and achieves good segmentation results which respect the realistic targets. This point can be proved by the fact that the proposed method outperformed the other ablation models, i.e., AtlasNet, AtlasNet-I, the Proposed method-LI, the Proposed method-I, the Proposed method-D1, and the Proposed method-D2.

4.6. Analysis of the training set size

The size of the training set has an important effect on the segmentation results of deep learning methods because deep learning methods generally require a large number of training data to learn plenty of parameters, which are used for extracting expressive features and to avoid over-fitting. However, even with a small training set the proposed method achieves outstanding performance, which has been shown in Fig. 7. Compared with the performance of 3D CNN (Milletari et al., 2016), with a decreased training set size, the performance of the proposed method is preferable even with slight reduction; but, the performance of 3D CNN decreases greatly. This is due to the proposed method explicitly integrating powerful atlas prior knowledge, which accurately and completely provides anatomical structure knowledge of the 3D LV. Based on the provided anatomical structure knowledge, the proposed method achieves segmentation only by learning specific geometrical deformation parameters (used for describing the variance between different objects), which can be learned by a light-weight network. Additionally, the parameters can be trained by a small training
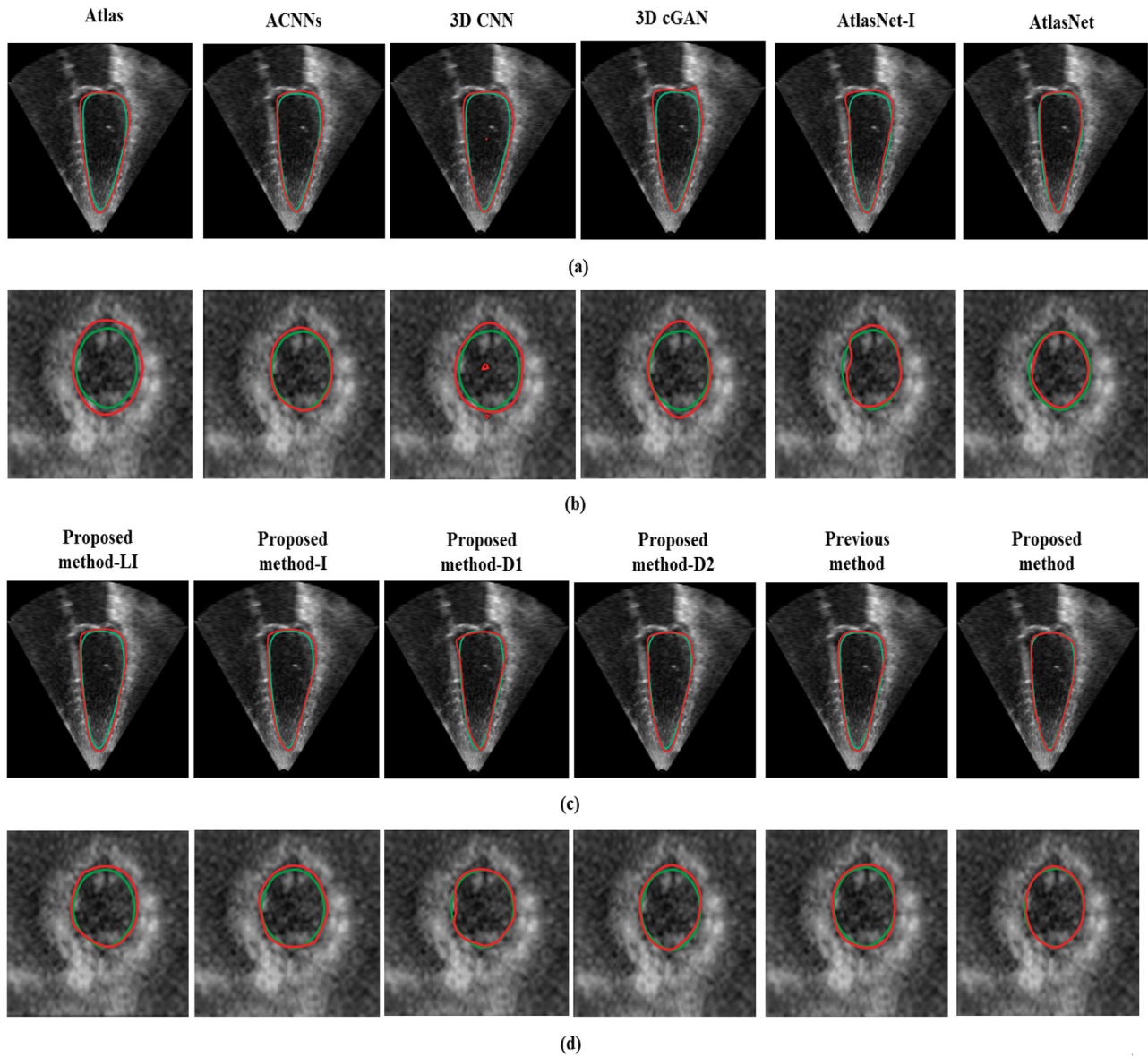
Fig. 8. The detailed comparison between the ground truth and the segmentation results of different methods in the 2D slices from a 3D volume. A green contour corresponds to the reference and a red contour corresponds to the result of the different methods.
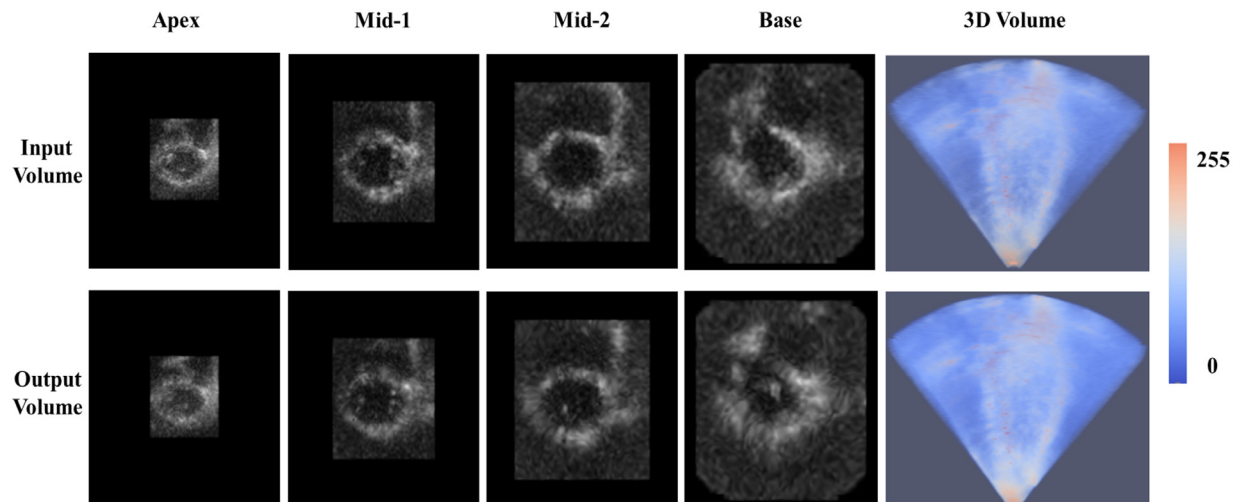

Fig. 9. The detailed comparison between the input volume and the output volume of the proposed method in the 2D slices from a 3D volume.

Table 4

The computation efficiency of different methods in the application and training stage.

Methods	FLOPs number for inference	Inference time(s)	Training time(h)	Number of parameters
3D Atlas	$4.63 \cdot 10^{13}$	4.18	N/A	N/A
3D CNN	$2.65 \cdot 10^{13}$	2.4	48	65.5M
3D cGAN	$2.65 \cdot 10^{13}$	2.4	60	66.7M
AtlasNet	$2.47 \cdot 10^{11}$	0.02	17	2.0M
Proposed method-LI	$2.47 \cdot 10^{11}$	0.02	24	3.2M
Proposed method	$2.47 \cdot 10^{11}$	0.02	30	3.2M

set. Hence, the proposed method achieves excellent performance with limited training data. It also means that the proposed method overcomes the problem of limited annotation data to some extent. However, having enough training data is important for the success of most deep learning methods, such as 3D CNN and the proposed method. Hence, collecting enough training data will be an important and long-time task in the 3DE LV segmentation field in the future. For future work, we will collect large-scale training data to explore new methods to address the LV segmentation problem for 3DE in large-scale training data. In the current study condition, the limited annotation data on 3DE is also an important challenge for LV segmentation. How to achieve good performance based on having limited data is an important research problem in the 3DE LV segmentation field inherently. The good segmentation performance in this current study condition demonstrates the potential of the proposed method for future clinical applications.

4.7. Analysis of inference efficiency

Inference efficiency is also very important for segmentation tasks, especially for medical image segmentation which is expected to be in real-time. In this paper, we focused on the difference in the number of floating point operations (FLOPs) and the inference time between our algorithm and the compared methods in order to make a more precise assessment of efficiency. Based on the widely used computation standard in (Pavlo et al., 2017) and algorithm flow in (Oktay et al., 2014), we computed the FLOPs of CNN-based methods and atlas-based methods. The computation efficiency has been shown in Table 4, whereby it requires $2.47 \cdot 10^{11}$ FLOPs (0.02s) which is lower than the compared methods. (1) The computation efficiency of the proposed method is 10^2 times higher than the 3D atlas method (Oktay et al., 2014) (The inference time of the proposed method is 209 times lower than the 3D atlas method), because the proposed method is based on an end-to-end off-line optimization and on-line inference. Hence, in the inference stage, the proposed method does not need iterative optimization like (Oktay et al., 2014) and is efficient. (2) The computation efficiency of the proposed method is 10^2 times higher than 3D CNN (Milletari et al., 2016) (The inference time of the proposed method is 120 times lower than the 3D CNN and 3D cGAN) and 3D cGAN. This is due to the proposed method being light-weight (The lower number of parameters is shown in Table 4) compared to the 3D CNN (Milletari et al., 2016) and 3D cGAN. Based on the integrated powerful atlas prior knowledge, the proposed method only needs to learn some specific knowledge to achieve segmentation. Moreover, the proposed method is able to achieve good performance from the multi-level information consistency constraint. Hence, compared with 3D CNN (Milletari et al., 2016) and 3D cGAN, the proposed method only used a light-weight network and achieved the best segmentation results. However, 3D CNN (Milletari et al., 2016) and 3D cGAN need to learn more parameters to acquire powerful information to achieve segmentation. This makes 3D CNN (Milletari et al., 2016) and 3D cGAN heavy-weight and inefficient. In contrast, the proposed method achieved

efficient inference even with the high dimensional data input of the 3DE volume.

5. Conclusion

In this paper, we proposed a deep atlas network with information consistency constraint for efficient 3D LV segmentation on 3DE. For the first time, the proposed method advantageously integrates atlas prior knowledge into a deep learning framework for 3D LV segmentation on echocardiography, which exploited a light-weight network to achieve better segmentation and efficient inference. Besides, the proposed framework can be trained with limited annotation data. Moreover, the proposed method uses the novel multi-level information consistency constraint (label consistency constraint, volume consistency constraint and adversarial consistency constraint) capable of enhancing the model's performance from different levels simultaneously and finally achieves 3D LV segmentation in a complex anatomical environment. The experimental results have demonstrated that the proposed method has clinical application potential for 3D LV segmentation on 3DE. Furthermore, we believe that the proposed method can also be used for other medical imaging tasks, which can be processed by atlas-based methods even though the annotation data is limited.

6. Funding information

Declaration of Competing Interest

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.

Acknowlgedgments

References

Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V., 2019. Voxelmorph: a learning framework for deformable medical image registration. IEEE transactions on medical imaging.

Barbosa, D., Dietenbeck, T., Heyde, B., Houle, H., Friboulet, D., Dhooge, J., Bernard, O., 2013. Fast and fully automatic 3-d echocardiographic segmentation using b-spline explicit active surfaces: feasibility study and validation in a clinical setting. Ultrasound in medicine & biology 39 (1), 89–101.

Barbosa, D., Heyde, B., Cikes, M., Dietenbeck, T., Claus, P., Friboulet, D., Bernard, O., Dhooge, J., 2014. Real-time 3d interactive segmentation of echocardiographic data through user-based deformation of b-spline explicit active surfaces. Computerized medical imaging and graphics 38 (1), 57–67.

Bernard, O., Bosch, J.G., Heyde, B., Alessandrini, M., Barbosa, D., Camarasu-Pop, S., Cervenansky, F., Valette, S., Mirea, O., Bernier, M., et al., 2016. Standardized evaluation system for left ventricular segmentation algorithms in 3d echocardiography. IEEE transactions on medical imaging 35 (4), 967–977.

Bernier, M., Jodoin, P.-M., Humbert, O., Lalande, A., 2017. Graph cut-based method for segmenting the left ventricle from mri or echocardiographic images. Computerized Medical Imaging and Graphics 58, 1–12.

Cao, X., Yang, J., Zhang, J., Wang, Q., Yap, P.-T., Shen, D., 2018. Deformable image registration using a cue-aware deep regression network. IEEE Transactions on Biomedical Engineering 65 (9), 1900–1911.

Carneiro, G., Nascimento, J.C., 2010. Multiple dynamic models for tracking the left ventricle of the heart from ultrasound data using particle filters and deep learning architectures. In: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, pp. 2815–2822.

Carneiro, G., Nascimento, J.C., 2013. Combining multiple dynamic models and deep learning architectures for tracking the left ventricle endocardium in ultrasound data. IEEE transactions on pattern analysis and machine intelligence 99 (1), 1.

Carneiro, G., Nascimento, J.C., Freitas, A., 2012. The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods. IEEE Transactions on Image Processing 21 (3), 968–982.

Chacko, R., Ramlingam Rajkumar, E., 2017. Three dimensional echocardiography: Recent trends in segmen tation methods. Current Medical Imaging Reviews 13 (3), 245–250.

Coppini, G., Poli, R., Valli, G., 1995. Recovery of the 3-d shape of the left ventricle from echocardiographic images. IEEE Transactions on Medical Imaging 14 (2), 301–317.

Corsi, C., Lang, R.M., Veronesi, F., Weinert, L., Caiani, E.G., MacEneaney, P., Lamberti, C., Mor-Avi, V., 2005. Volumetric quantification of global and regional left ventricular function from real-time three-dimensional echocardiographic images. Circulation 112 (8), 1161–1170.

Dong, S., Luo, G., Sun, G., Wang, K., Zhang, H., 2016. A left ventricular segmentation method on 3d echocardiography using deep learning and snake. In: Computing in Cardiology Conference (CinC), 2016. IEEE, pp. 473–476.

Dong, S., Luo, G., Wang, K., Cao, S., Mercado, A., Shmuilovich, O., Zhang, H., Li, S., 2018. Voxelatlasgan: 3d left ventricle segmentation on echocardiography with atlas guided generation and voxel-to-voxel discrimination. MICCAI.

Fan, J., Cao, X., Yap, P.-T., Shen, D., 2019. Birnet: Brain image registration using dual–supervised fully convolutional networks. Medical image analysis 54, 193–206.

Ge, R., Yang, G., Chen, Y., Luo, L., Feng, C., Ma, H., Ren, J., Li, S., 2019. K-net: Integrate left ventricle segmentation and direct quantification of paired echo sequence. IEEE transactions on medical imaging.

Ge, R., Yang, G., Chen, Y., Luo, L., Feng, C., Zhang, H., Li, S., 2019. Pv-lvnet: Direct left ventricle multitype indices estimation from 2d echocardiograms of paired apical views with deep neural networks. Medical image analysis 58, 101554.

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.

Hansegard, J., Urheim, S., Lunde, K., Rabben, S.I., 2007. Constrained active appearance models for segmentation of triplane echocardiograms. IEEE transactions on medical imaging 26 (10), 1391–1400.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision, pp. 1026–1034.

Hu, Y., Modat, M., Gibson, E., Li, W., Ghavami, N., Bonmati, E., Wang, G., Bandula, S., Moore, C.M., Emberton, M., et al., 2018. Weakly-supervised convolutional neural networks for multimodal image registration. Medical image analysis 49, 1–13.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: CVPR, 1, pp. 4700–4708.

Huang, X., Dione, D.P., Compas, C.B., Papademetris, X., Lin, B.A., Bregasi, A., Sinusas, A.J., Staib, L.H., Duncan, J.S., 2014. Contour tracking in echocardiographic sequences via sparse representation and dictionary learning. Medical image analysis 18 (2), 253–271.

Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167.

Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 5967–5976.

Jaderberg, M., Simonyan, K., Zisserman, A., et al., 2015. Spatial transformer networks. In: Advances in neural information processing systems, pp. 2017–2025.

Juang, R., McVeigh, E.R., Hoffmann, B., Yuh, D.D., Burlina, P., 2011. Automatic segmentation of the left-ventricular cavity and atrium in 3d ultrasound using graph cuts and the radial symmetry transform. In: ISBI. Citeseer, pp. 606–609.

Krebs, J., e Delingette, H., Mailhé, B., Ayache, N., Mansi, T., 2019. Learning a probabilistic model for diffeomorphic registration. IEEE transactions on medical imaging.

Krishnaswamy, D., Hareendranathan, A.R., Suwatanaviroj, T., Becher, H., Noga, M., Punithakumar, K., 2018. A semi-automated method for measurement of left ventricular volumes in 3d echocardiography. IEEE Access 6, 16336–16344.

Lang, R.M., Badano, L.P., Tsang, W., Adams, D.H., Agricola, E., Buck, T., Faletra, F.F., Franke, A., Hung, J., de Isla, L.P., et al., 2012. Eae/ase recommendations for image acquisition and display using three-dimensional echocardiography. European Heart Journal–Cardiovascular Imaging 13 (1), 1–46.

Lempitsky, V., Verhoek, M., Noble, J.A., Blake, A., 2009. Random forest classification for automatic delineation of myocardium in real-time 3d echocardiography. In: International Conference on Functional Imaging and Modeling of the Heart. Springer, pp. 447–456.

Leung, K.E., Bosch, J.G., 2010. Automated border detection in three-dimensional echocardiography: principles and promises. European journal of echocardiography 11 (2), 97–108.

Luo, G., Dong, S., Wang, K., Zuo, W., Cao, S., Zhang, H., 2018. Multi-views fusion cnn for left ventricular volumes estimation on cardiac mr images. IEEE Transactions on Biomedical Engineering 65 (9), 1924–1934.

Luo, G., Dong, S., Wang, W., Wang, K., Cao, S., Tam, C., Zhang, H., Howey, J., Ohorodnyk, P., Li, S., 2020. Commensal correlation network between segmentation and direct area estimation for bi-ventricle quantification. Medical image analysis 59, 101591.

Ma, M., van Stralen, M., Reiber, J.H., Bosch, J.G., Lelieveldt, B.P., 2010. Model driven quantification of left ventricular function from sparse single-beat 3d echocardiography. Medical image analysis 14 (4), 582–593.

Mellor, M., Brady, M., 2005. Phase mutual information as a similarity measure for registration. Medical image analysis 9 (4), 330–343.

Miao, S., Wang, Z.J., Liao, R., 2016. A cnn regression approach for real-time 2d/3d registration. IEEE transactions on medical imaging 35 (5), 1352–1363.

Milletari, F., Navab, N., Ahmadi, S.-A., 2016. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 3D Vision (3DV), 2016 Fourth International Conference on. IEEE, pp. 565–571.

Milletari, F., Yigitsoy, M., Navab, N., 2014. Left ventricle segmentation in cardiac ultrasound using hough-forests with implicit shape and appearance priors. midas.

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

Mitchell, S.C., Bosch, J.G., Lelieveldt, B.P., Van der Geest, R.J., Reiber, J.H., Sonka, M., 2002. 3-d active appearance models: segmentation of cardiac mr and ultrasound images. IEEE transactions on medical imaging 21 (9), 1167–1178.

Mo, Y., Liu, F., McIlwraith, D., Yang, G., Zhang, J., He, T., Guo, Y., 2018. The deep poincaré map: A novel approach for left ventricle segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 561–568.

Mukherjee, R., Vyas, S., Juang, R., Sprouse, C., Burlina, P., 2013. Endocardial surface delineation in 3-d transesophageal echocardiography. Ultrasound in medicine & biology 39 (12), 2447–2462.

Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M., Bai, W., Caballero, J., Cook, S.A., de Marvao, A., Dawes, T., ORegan, D.P., et al., 2018. Anatomically constrained neural networks (acnns): application to cardiac image enhancement and segmentation. IEEE transactions on medical imaging 37 (2), 384–395.

Oktay, O., Shi, W., Keraudren, K., Caballero, J., Rueckert, D., Hajnal, J., 2014. Learning shape representations for multi-atlas endocardium segmentation in 3d echo images. Proceedings MICCAI Challenge on Echocardiographic Three-Dimensional Ultrasound Segmentation (CETUS), Boston, MIDAS Journal 57–64.

Orderud, F., Rabben, S.I., 2008. Real-time 3d segmentation of the left ventricle using deformable subdivision surfaces. In: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference On. IEEE, pp. 1–8.

Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., Efros, A.A., 2016. Context encoders: Feature learning by inpainting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2536–2544.

Pavlo, M., Stephen, T., Tero, K., Timo, A., Jan, K., 2017. Pruning convolutional neural networks for resource efficient inference. In: International Conference on Learning Representations, pp. 1–17.

Pedrosa, J., Barbosa, D., Almeida, N., Bernard, O., Bosch, J., et al., 2016. Cardiac chamber volumetric assessment using 3d ultrasound-a review. Current pharmaceutical design 22 (1), 105–121.

Pedrosa, J., Queirós, S., Bernard, O., Engvall, J., Edvardsen, T., Nagel, E., Dhooge, J., 2017. Fast and fully automatic left ventricular segmentation and tracking in echocardiography using shape-based b-spline explicit active surfaces. IEEE transactions on medical imaging 36 (11), 2287–2296.

Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O., Hawkes, D.J., 1999. Nonrigid registration using free-form deformations: application to breast mr images. IEEE transactions on medical imaging 18 (8), 712–721.

Santiago, C., Nascimento, J.C., Marques, J.S., 2015. Automatic 3-d segmentation of endocardial border of the left ventricle from ultrasound images. IEEE journal of biomedical and health informatics 19 (1), 339–348.

Schneider, R.J., Perrin, D.P., Vasilyev, N.V., Marx, G.R., Pedro, J., Howe, R.D., 2012. Real-time image-based rigid registration of three-dimensional ultrasound. Medical image analysis 16 (2), 402–414.

Smistad, E., Lindseth, F., 2014. Real-time tracking of the left ventricle in 3d ultrasound using kalman filter and mean value coordinates. In: Proceedings of MICCAI Challenge on Echocardiographic Three-Dimensional Ultrasound Segmentation (CETUS). MIDAS Journal, pp. 65–72.

Van Assen, H.C., Danilouchkine, M.G., Frangi, A.F., Ordás, S., Westenberg, J.J., Reiber, J.H., Lelieveldt, B.P., 2006. Spasm: a 3d-asm for segmentation of sparse and arbitrarily oriented cardiac mri data. Medical Image Analysis 10 (2), 286–303.

de Vos, B.D., Berendsen, F.F., Viergever, M.A., Sokooti, H., Staring, M., Išgum, I., 2019. A deep learning framework for unsupervised affine and deformable image registration. Medical image analysis 52, 128–143.

Wu, F., Yang, X.H., Packard, A., Becker, G., 1996. Induced l2-norm control for lpv systems with bounded parameter variation rates. International Journal of Robust and Nonlinear Control 6 (9-10), 983–998.

Xue, W., Islam, A., Bhaduri, M., Li, S., 2017. Direct multitype cardiac indices estimation via joint representation and regression learning. IEEE transactions on medical imaging 36 (10), 2057–2067.

Xue, W., Lum, A., Mercado, A., Landis, M., Warrington, J., Li, S., 2017. Full quantification of left ventricle via deep multitask learning network respecting intra-and inter-task relatedness. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 276–284.

Xue, W., Nachum, I.B., Pandey, S., Warrington, J., Leung, S., Li, S., 2017. Direct estimation of regional wall thicknesses via residual recurrent neural network. In: International Conference on Information Processing in Medical Imaging. Springer, pp. 505–516.

Yang, L., Georgescu, B., Zheng, Y., Wang, Y., Meer, P., Comaniciu, D., 2011. Prediction based collaborative trackers (pct): A robust and accurate approach toward 3d medical object tracking. IEEE Transactions on Medical Imaging 30 (11), 1921–1932.

Yu, Y., Zhang, S., Li, K., Metaxas, D., Axel, L., 2014. Deformable models with sparsity constraints for cardiac motion analysis. Medical image analysis 18 (6), 927–937.

Zagrodsky, V., Walimbe, V., Castro-Pareja, C.R., Qin, J.X., Song, J.-M., Shekhar, R., 2005. Registration-assisted segmentation of real-time 3-d echocardiographic data using deformable models. IEEE Transactions on Medical Imaging 24 (9), 1089–1099.

Zhang, H., He, X., 2017. Deep free-form deformation network for object-mask registration. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 4251–4259.

Zhang, S., Zhan, Y., Dewan, M., Huang, J., Metaxas, D.N., Zhou, X.S., 2012. Towards robust and effective shape modeling: Sparse shape composition. Medical image analysis 16 (1), 265–277.

Zhang, W., Noble, J.A., Brady, J.M., 2007. Adaptive non-rigid registration of real time 3d ultrasound to cardiovascular mr images. In: Biennial International Conference on Information Processing in Medical Imaging. Springer, pp. 50–61.

Zhen, X., Wang, Z., Yu, M., Li, S., 2015. Supervised descriptor learning for multi-output regression. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, pp. 1211–1218.

Zhu, Y., Papademetris, X., Sinusas, A.J., Duncan, J.S., 2010. A coupled deformable model for tracking myocardial borders from real-time echocardiography using an incompressibility constraint. Medical Image Analysis 14 (3), 429–448.

Zhuang, X., Rhode, K.S., Razavi, R.S., Hawkes, D.J., Ourselin, S., 2010. A registration-based propagation framework for automatic whole heart segmentation of cardiac mri. IEEE transactions on medical imaging 29 (9), 1612–1625.

Zhuang, X., Yao, C., Ma, Y., Hawkes, D., Penney, G., Ourselin, S., 2010. Registration-based propagation for whole heart segmentation from compounded 3d echocardiography. In: Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on. IEEE, pp. 1093–1096.