

# Anatomically-guided deep learning for left ventricle geometry generation with uncertainty quantification based on short-axis MR images

Andre Von Zuben<sup>a</sup>, Luigi E. Perotti<sup>a</sup> and Felipe A. C. Viana<sup>a,\*</sup>

<sup>a</sup>Department of Mechanical and Aerospace Engineering, University of Central Florida, Orlando, FL 32816-8030, USA

## ARTICLE INFO

### Keywords:

Anatomically-guided deep learning  
MRI-based geometry generation  
Applied machine learning  
B-spline head

## ABSTRACT

Recent advances in deep learning have greatly improved the ability to generate analysis models from medical images. In particular, great attention is focused on quickly generating models of the left ventricle from cardiac magnetic resonance imaging (cMRI) to improve the diagnosis and prognosis of millions of patients. However, even state-of-the-art frameworks present challenges, such as discontinuities of the cardiac tissue and excessive jaggedness along the myocardial walls. These geometrical features are often anatomically incorrect and may lead to unrealistic results once the geometrical models are employed in computational analyses. In this work, we propose an anatomically-guided deep learning model to overcome these limitations while preserving the advantages of state-of-the-art frameworks, such as computational efficiency, robustness, and generalization capabilities. Our novel anatomically-guided neural networks are formed by a UNet followed by a B-spline head, which acts as a regularization layer during training. The B-spline head aggregates the prediction into a single connected region, removes any undesired tissue islands, and produces a smooth continuous contour. In addition, the introduction of the B-spline head contributes to achieve a robust uncertainty quantification of the left ventricle inner and outer walls. Our results show that the proposed model generates anatomically consistent geometries while achieving an agreement with the ground truth images comparable to state-of-the-art frameworks and simultaneously improving the geometry uncertainty quantification in comparison to classic UNet models. The examples presented here, as well as source codes, are all open-source under the GitHub repository [https://github.com/CBL-UCF/unet\\_ag](https://github.com/CBL-UCF/unet_ag).

## 1. Introduction

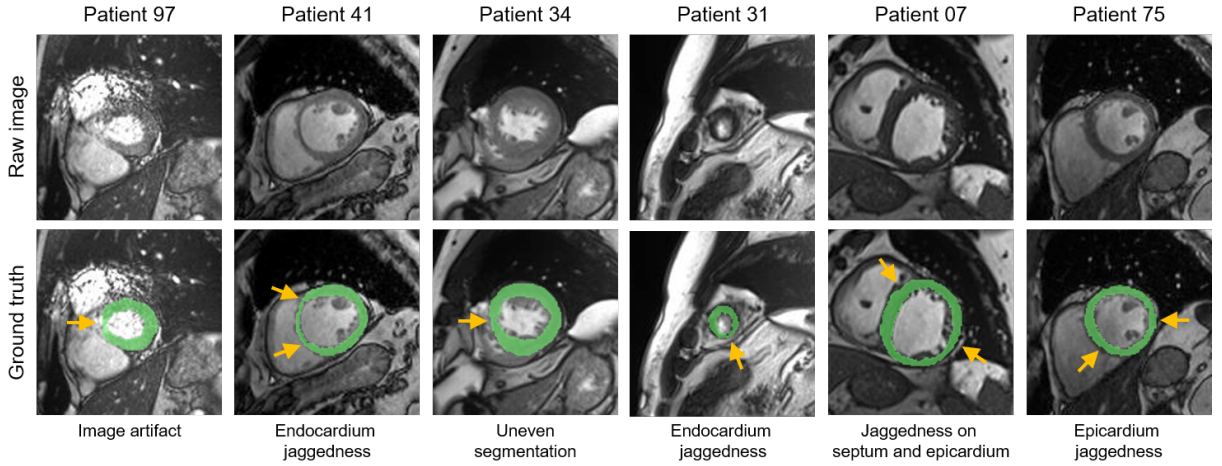
Computational cardiology has shown large potential to improve the diagnosis, prognosis, and therapy planning for patients affected by cardiomyopathies [1]. Patient-specific analyses of cardiac function and dysfunction require models that accurately represent the patient's cardiac anatomy. For example, image-based cardiac models are often adopted to study cardiac kinematics (e.g., [2, 3]) and mechanics (e.g., [4, 5]) with the goal of evaluating cardiac function and dysfunction. The necessary geometrical models can be generated from magnetic resonance imaging (MRI) by manually segmenting the myocardial tissue. For example, the geometry of the left ventricle can be reconstructed by segmenting short-axis images (acquired perpendicular to the ventricle longitudinal axis) and combining in 3D the resulting annular regions. However, generating patient-specific models by manually segmenting imaging data is labor intensive, error prone, and introduces observer variability. These factors limit the feasibility and robustness of patient-specific cardiac models, and therefore hinder their large-scale clinical adoption.

In order to overcome the obstacles due to manual segmentation, in recent years, machine learning has been deployed to segment imaging data and create patient-specific geometries. Despite the advances and growing adoption of machine learning based segmentations in cardiac analyses, key problems still exist with the current predictions, e.g., discontinuities in the cardiac tissue, jaggedness of the endocardial and epicardial surfaces, and over-prediction of the myocardium at the right ventricle insertion points, which leads to anatomically incorrect features in the resulting segmentation [6, 7]. These anatomically incorrect features can be caused by low contrast or artifacts in the acquired images (leading to discontinuity in the predicted myocardium) and limitations in the image resolution (leading to jagged segmentations). In addition, limitations in image contrast and resolution may affect the creation of ground

\*Corresponding author

 avzuben@knights.ucf.edu (A.V. Zuben); Luigi.Perotti@ucf.edu (L.E. Perotti); <https://pml-ucf.github.io/>, viana@ucf.edu (F.A.C. Viana)

ORCID(s): 0000-0003-4433-2608 (A.V. Zuben); 0000-0002-9010-2144 (L.E. Perotti); 0000-0002-2196-7603 (F.A.C. Viana)



**Figure 1:** Examples of typical challenges encountered with cardiac magnetic resonance (cMR) images and their ground truth segmentations. The presented imaging data is part of the ACDC challenge dataset provided in [8] and associated with [9].

truth labels, which are further compounded by observer variability (e.g., different observers may unevenly segment the myocardium near trabeculae and papillary muscles). Examples of these challenges are illustrated in Fig. 1, which is based on the data shared by the Automated Cardiac Diagnosis Challenge (ACDC) [8, 9]. The ACDC dataset has been used throughout this study.

An additional key aspect in constructing anatomically accurate geometries is the ability to quantify the model uncertainty reflecting, for example, the image quality. This is necessary to associate the patient-specific models with a confidence level for the clinical predictions deriving from the use of the models. Segmentation approaches based on deep learning can assign probabilities for specific features, such as the left ventricle cavity or the left ventricle myocardium. However, these uncertainty estimates are neither optimized nor robust.

In this context, the key motivation of this work is to overcome the aforementioned drawbacks present in current machine-learning based segmentations and generate geometrical models for high-fidelity, physics-based simulations, such as the ones based on finite element analyses. This requires: (1) anatomically accurate geometry generation without discontinuity, jaggedness, and over- or under-prediction due to image resolution; and (2) robust uncertainty quantification in the predicted geometries. We achieve these goals by designing and implementing a pipeline of deep learning models composed of a left ventricle locator and two *anatomically-guided deep neural networks*. These anatomically-guided deep neural networks are the main contribution of our work. Their architecture consists in a UNet followed by a specialized B-spline head, which we introduce in this paper. The B-spline head greatly improves the UNets training and provide the desired degree of anatomical fidelity in their final prediction. As opposed to a simple post-processing step, the B-spline head is used during training of the UNets and works by constraining their predictions, eliminating islands and ensuring smooth contours.

The remaining of this manuscript is organized as follows. Section 2 contextualizes this work in the current literature by outlining several key examples using deep learning for image segmentation and geometry generation. Section 3 describes the proposed method, including the left ventricle cavity locator, the anatomically-guided deep learning models, and the model uncertainty quantification. Section 4 details the case study adopted in this work, the training of the proposed model, and how to replicate the presented results. Finally, Section 5 outlines and discusses the results obtained with the newly proposed model, while Section 6 recapitulates the key points of this work and presents conclusions and future work.

## 2. Background and previous work

Amongst the many contributions in the field of medical image segmentations using machine learning, here we introduce a few references to contextualize the current work in literature. These works show alternatives approaches to the one proposed here, whose main objectives are to produce an anatomically consistent geometry with robust

uncertainty quantification.

*Deep learning for medical imaging* - In recent years, deep learning techniques have been extensively used for medical image segmentation and analysis. For example, deep learning approaches have been widely deployed in cardiac image analysis, especially MR images, where state-of-the-art methods can successfully obtain precise segmentations and categorize patient data [9–11]. One of the main factors responsible for these advances was the introduction of the UNet by Ronneberger et al. [12], which achieved remarkable performance in many diverse image segmentations in medical applications. The UNet consists in an encoder followed by a decoder sub-network, with several skipped connections between the down and upsampling paths. Zhou et al. [13] proposed the UNet++ architecture as an expansion of the UNet. In this new approach, the encoder and decoder sub-networks are connected through nested dense skip pathways, aiming to decrease the semantic gap between the encoder’s and decoder’s feature maps. More recently, Li et al. [14] introduced a cascade UNet model for prostate segmentation consisting of two stages: the first stage provides preliminary segmentation results that are then utilized as the input for the second stage. Furthermore, Yang et al. [15] proposed a fully automatic framework based on a customized 3D fully convolutional segmentation network (FCN) for ventricular segmentation in MRI volumes, which uses a class-balanced loss for aiding the optimization. Finally, Baumgartner et al. [16] investigated the performance of different architectures, namely an FCN, a 2D-UNet, a modified 2D-UNet, and a modified 3D-UNet, for segmentation of cardiac MRI, and showed that the 2D-UNets overperformed the other approaches.

*Identifying the region of interest (ROI)* - Oktay et al. [17], inspired by [18], proposed an Attention UNet composed of grid-based gating that permits attention coefficients to be more localized. Li et al. [19] designed a multi-stage architecture and attention blocks to overcome the challenges brought by the segmentation of small regions. Islam et al. [20] presented a 3D-UNet architecture and integrated channel and spatial attention with a decoder sub-network for brain tumor segmentation. Likewise, Cheng et al. [21] proposed a triple attention block for 3D point cloud semantic segmentation to provide better context information, which comprises position, channel, and local region attention modules. Instead of using attention approaches, Ferdian et al. [22] employed a localization convolutional neural network that outputs a bounding box around the ROI, which is then used for cropping the image and removing any background noise. Alternatively, Khened et al. [23] extracted ROIs based on the spatio-temporal statistical analysis of the segmented cardiac images and the circle Hough Transform technique to separate and outline the heart structures from the adjacent tissues. Finally, Zhao et al. [24] introduced a cascaded UNet-like model to aid the lung tumor segmentation task. The first stage of the network delivers a global prediction that provides context information used by a distraction-attention module, which guides the second stage of the model and produces a refined segmentation.

*Simultaneous use of multiple deep neural networks* - Isensee et al. [25] combined 2D- and 3D-UNets predictions using an arithmetic average to improve the final segmentation. Recently, they expanded their work proposing the nnUNet [26], an automated multi-domain medical imaging segmentation framework that was applied to multiple image modalities (e.g., MRI, computed tomography) and datasets outperforming most of the state-of-the-art pipelines. This framework includes a 2D, a 3D, and a cascaded 3D-UNet. In our previous work [27], we have also presented a pipeline composed of an ROI Locator UNet, a 2D-UNet, and a 2D-UNet++. The models’ results were combined using weights inversely proportional to their mean loss.

*Geometry-constrained cardiac image segmentation* - Popescu et al. [28] proposed an “anatomically informed deep learning” approach to analyze contrast-enhanced cMRI for scar segmentation. This pipeline consists of 3 sub-networks: an ROI Locator UNet, a left ventricle (LV) Segmentation UNet, and an autoencoder UNet responsible for encoding and decoding myocardial segmentation masks. The segmentation labels are encoded using the third sub-network to create a latent space. Predicted segmentations are also encoded using this sub-network, and the nearest-neighbors algorithm is used to generate an alternative anatomically correct version of the initial segmentation. Grinias and Tziritas [29] used B-splines to smooth the LV epicardium predictions as a post-processing step. The predictions are obtained from a fast fully automatic method based on a Markov random field (MRF) model.

*Heart mesh reconstruction* - In recent years, there have been several studies regarding the generation of cardiac meshes based on MRI segmentation [30, 31] or volumetric CT and MR data [32], with the goal of avoiding common problems such as discontinuities and jaggedness. Beetz et al. [30] presented a deep learning technique based on point completion networks capable of generating biventricular 3D surface reconstructions from cardiac MRI outlines in a

rapid and fully automated process. Banerjee et al. [31] presented a pipeline for generating patient-specific 3D biventricular heart models from MRI. Their framework chooses the relevant MR images, segments them, and extracts the heart contours. The contours are then aligned in 3D space using the intensity and the contours information from cine data aided by a statistical shape model. Subsequently, the 3D representation of the contours is used to generate a smooth 3D biventricular mesh. Kong et al. [32] proposed a deep-learning-based approach for mesh reconstruction of the entire heart from volumetric CT and MR data by using a graph convolutional neural network that deforms the vertices from a predefined mesh template in order to accommodate the anatomical structures present in the image volume.

In this paper, we propose a pipeline composed of three deep neural network models. The first network is used to isolate the ROI – the left ventricle in our case. Then two modified UNets generate the inner and outer wall of the left ventricle. The UNets are augmented by B-spline heads to form the anatomically-guided deep neural networks. By incorporating a B-spline head as a regularization layer, the UNets exhibit the B-spline’s smoothing and continuity effect. The choice of including a B-spline head was further motivated by our previous work [27], in which B-splines were employed for contour extraction as an intermediate step between segmentation and mesh generation.

### 3. Proposed geometry generation pipeline and anatomically-guided deep learning

Three neural networks form the proposed pipeline: a LVC-locator and two UNet<sub>AG</sub> for LVC and LVM predictions (Fig. 2). Their two fundamental tasks are the:

1. Localization of the left ventricular cavity. The raw MR input images are fed to the first deep neural network (LVC-Locator) to find the region of interest, i.e., the left ventricular cavity. A UNet is adopted to achieve this purpose. In practical terms, by cropping the image around the region of interest, the LVC-Locator allows for a targeted feature extraction in the next step.
2. Anatomically-guided feature extraction. The output of the LVC-Locator is used to feed two anatomically-guided deep neural networks, which will produce the geometrical contours of the left ventricle cavity (LVC) and left ventricle myocardium (LVM), i.e., the LV inner and outer walls. Fig. 2 details the architectures of the LVC-Locator as well as the LVC- and LVM-UNet<sub>AG</sub> models. The first anatomically-guided deep neural networks is the LVC-UNet<sub>AG</sub>, which takes the cropped image as input and returns the contour of the left ventricle inner (endocardial) wall. Similarly, the second deep neural network is the LVM-UNet<sub>AG</sub>, which takes the cropped image alongside the LVC-UNet<sub>AG</sub> prediction as inputs and returns the contour of the left ventricle outer (epicardial) wall.

Both anatomically-guided neural networks contain two important elements: a UNet for segmentation and a B-spline head for conditioning of training. As we will further detail in the next section, the B-spline head acts as a constraint during training such that predictions resulting in smooth and continuous contours are favored. This is a feature of B-splines [33], which are commonly used in computer-aided design for geometry generation. It is important to note that, as opposed to being used in a post-processing step, the B-splines are employed during training and their smoothing and continuity characteristics affect the stochastic gradient descent training of the neural network.

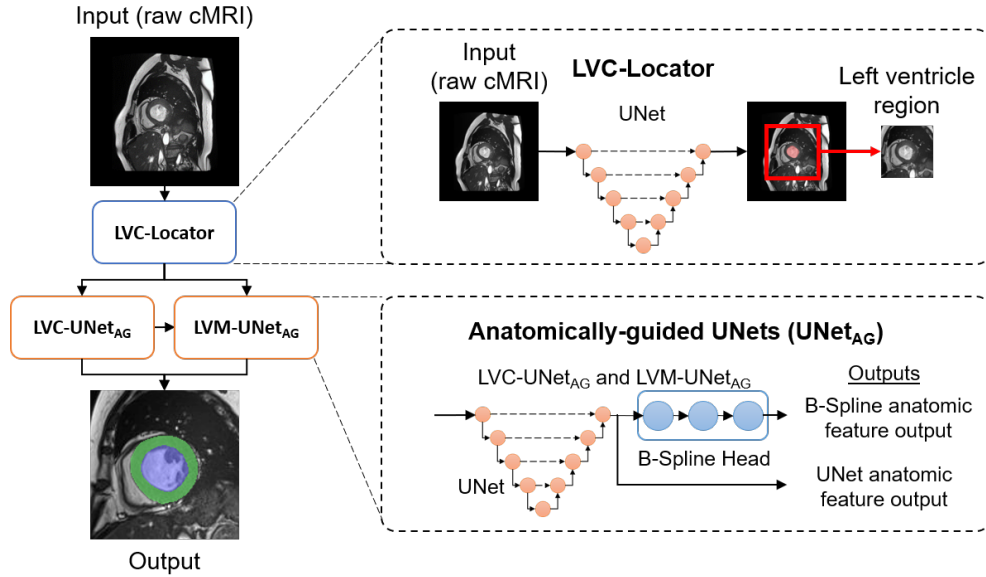
Next, we will describe in detail each component of the proposed pipeline, the associated loss functions, and the training process.

#### 3.1. Left ventricle cavity locator

A deep neural network delineates the region of interest before proceeding to generate the left ventricle geometry. This step reduces the overall image size provided to the UNet<sub>AG</sub> models while also centering the left ventricle in the input image. Similarly to our previous work [27], we use a UNet as left ventricle cavity locator (LVC-Locator). This network receives raw cMRI and outputs each pixel’s probability to be part of the LVC. In order to train the model, we use a compound loss function that is the summation of a Dice Similarity Coefficient (DSC) loss and a cross-entropy loss:

$$\mathcal{L}_{\text{LVC-Locator}} = -\frac{2}{|C|} \sum_{c \in C} \frac{\sum_i u_i^c l_i^c}{\sum_i u_i^c + \sum_i l_i^c} - \sum_{c \in C} \sum_i l_i^c \log(u_i^c), \quad (1)$$

where  $u$  is the output of the network,  $l$  is a one hot encoding of the label segmentation map,  $i$  is the voxel number in each map, and  $c \in C$  are the classes.  $u$  and  $l$  are of size  $N_i \times N_c$ , where  $N_i$  and  $N_c$  are the number of voxels in



**Figure 2:** Proposed pipeline of deep learning models for left ventricle geometry generation.

each map and the number of classes, respectively. Given that the main goal of the LVC-Locator is to localize the LVC region for the next two UNet<sub>AG</sub> networks in our pipeline, in Eq. (1) we used two classes only: LVC and background.

The LVC-locator is used to estimate the midpoint of the region of interest (i.e., the center of left ventricle cavity). First, a bounding box from the pixels predicted to contain the LVC (largest and smallest horizontal and vertical locations) is generated. Subsequently, the midpoint of the bounding box is computed. For every patient and cardiac phase, we pass the full short-axis images' stack to the LVC-Locator so that a single central point per patient and cardiac phase is identified by averaging the central point locations of all slices. This process has proven to increase robustness, especially in the bottom and top slices (i.e., the most apical and basal slices, respectively). Cropped images for each slice are generated by extracting a region of pre-defined size centered around the average midpoint. Both anatomically-guided neural networks will receive these cropped images as inputs.

### 3.2. Anatomically-guided deep learning models

In this paper, we introduce anatomically-guided neural networks formed by a UNet followed by a B-spline head (Fig. 2). The key components of these models are the:

- *UNet*. The UNet is composed of an encoder containing four convolutional blocks, each formed by two 2D convolutional layers and batch normalization, a feature space convolutional block, and four additional transposed convolutional blocks in the decoder, with skip connections between mirrored layers in the encoder and decoder stacks. We use the same architecture used in our previous work [27] following the work of Isensee et al. [25]. The UNet components outputs the probability of two different classes for each pixel: background and foreground (LVC).
- *B-spline head*. The expanded neural network architecture with the layers forming the B-spline head is the main innovation of our work. This network head is composed of three layers; namely a contour detection, a B-spline, and a perimeter-filling layers. Its task is to receive the outputs of the UNet and return a smooth and constrained region with high probability of containing the region of interest.
- *Left ventricle cavity and myocardium UNet<sub>AG</sub> models*. The proposed pipeline is comprised of two anatomically-guided models, one for the left ventricle cavity (LVC-UNet<sub>AG</sub>) and another for the left ventricle myocardium (LVM-UNet<sub>AG</sub>). While the LVC-UNet<sub>AG</sub> receives as input only the cropped image from the LVC-locator, the LVM-UNet<sub>AG</sub> also receives as inputs the predictions of the LVC-UNet<sub>AG</sub>. The staggered inputs improve the overall left ventricle geometry generated by the pipeline. In this work, the LVC-UNet<sub>AG</sub> and the LVM-UNet<sub>AG</sub>

architectures are identical except with regard to the following two aspects: (i) they output either the probability of belonging to the LVC or the LVM (in addition to the probability of belonging to the background); and (ii) the LVC-UNet<sub>AG</sub> receives the LVC-UNet<sub>AG</sub> output as an additional input channel. This additional input improves the learning of an anatomically feasible wall thickness in the LVM-UNet<sub>AG</sub>.

The B-spline head plays an essential part during training as it helps the LVC- and LVM-UNet<sub>AG</sub> by removing all the islands in the predictions, minimizing the spikes in the contours, and ensuring the continuity of the predicted region, therefore helping to guide the LVC- and LVM-UNet<sub>AG</sub> gradients. In other words, the B-spline head acts as a regularizer during training, aggregating the predictions into a single region, providing a smooth contour, and guaranteeing continuity. After training, only the UNet outputs are used for prediction. The final left ventricle geometry is generated by subtracting the LVC-UNet<sub>AG</sub> output from the LVM-UNet<sub>AG</sub> output.

### 3.2.1. B-spline head

Fig. 3 illustrates the proposed B-spline head. The main objective behind this neural network head is to regularize the training of the UNet such that anatomically inaccurate features are removed in the training process. Accordingly, the UNet training is focused on defining the region of interest while avoiding abnormalities such as jaggedness and islands. The B-spline head takes the segmentation from the UNet and creates its B-spline representation. In order to implement this aspect in a manner that is friendly to the stochastic gradient descent algorithm used in training (allowing backward propagation of gradients), the B-spline head is composed of three layers:

- *The contour layer*, which receives an input of shape  $(b, h, w)$ , where  $b$  is the batch size,  $h$  is the height of the prediction map, and  $w$  is the width of the prediction map generated by the UNet associated with the UNet<sub>AG</sub> model. The contour layer outputs  $n$  two-dimensional points representing the contour of the anatomic feature (here, we chose  $n = 360$  points spaced by  $1^\circ$  in polar coordinates). In terms of implementation we: (1) split the image into left and right sides with respect to the center point; (2) scan each row of the probability map from each side separately to find the left and right contours of the anatomical feature; and (3) sample the identified contour with the selected  $n$  points. If the UNet outputs an anatomically consistent shape without islands or ‘protrusions’, the contour layer returns only the outer perimeter of the UNet output (Fig. 3a). However, if unrealistic features that tend to be highly uncorrelated with the main region of interest are present, the contour layer will reduce their impact. This is demonstrated in Fig. 3b where the effect of the B-Spline head ameliorate the unrealistic spikes present in the first epoch. The unrealistic feature will then rapidly disappear in subsequent epochs.
- *The spline layer*, which receives as input the contours of the anatomic features generated by the contour layer and outputs their respective two-dimensional spline contours. In terms of implementation, the B-spline control points are a subset of the points that describe the contour of the anatomic feature. In this work, we select 20 control points per B-spline. The control points are equally spaced by  $18^\circ$  based on empirical analyses performed in our previous work [27]. The spline layer constrains the predicted anatomic feature to be described by a smooth and fully enclosed region. Constraining the predicted region contour with a B-spline drastically minimizes areas with spikes that can be anatomically unrealistic and would otherwise be mispredicted. While this layer has minimal effect (other than smoothing) when the anatomic features are realistic (Fig. 3a), it substantially improves contour predictions for unrealistic anatomic features (Fig. 3b).
- *The filling layer*, which receives any number of points in Cartesian coordinates representing a closed region and returns a  $h \times w$  matrix containing values from 0 to 1 (recall that the UNet<sub>AG</sub> models take  $h \times w$  images as inputs). Each position in the matrix will be 1 if its coordinates are enclosed by the anatomic feature or 0 otherwise. In practice, this layer returns the enclosure of the anatomic feature delineated by the B-spline contour.

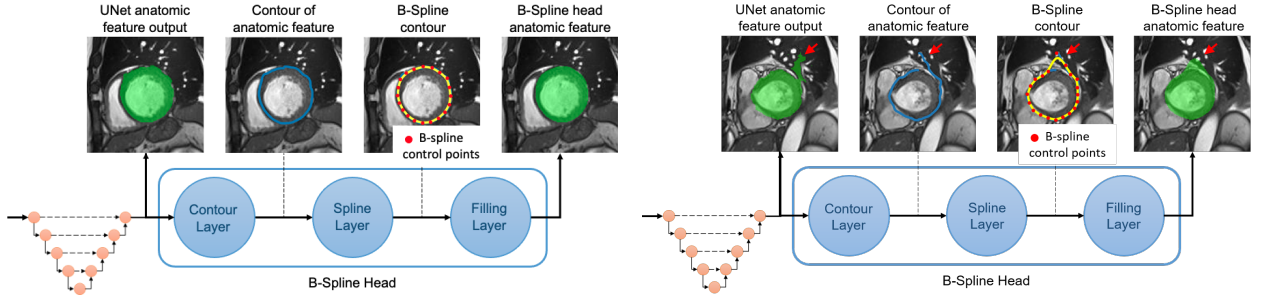
The proposed B-spline head generates a smooth and constrained anatomic feature without any parameter optimization, e.g., without optimizing the number of B-spline control points. Based on empirical studies, twenty control points are sufficient for the application reported here and no optimization was necessary in this case.

### 3.2.2. UNet<sub>AG</sub> training and loss functions

The most important aspect of the B-spline head is aiding the training of the LVC- and LVM-UNet<sub>AG</sub> models. As discussed before, the architectures of both LVC- and LVM-UNet<sub>AG</sub> models are very similar<sup>1</sup> and the only trainable

<sup>1</sup>The LVM-UNet<sub>AG</sub> receives the output of the LVC-UNet<sub>AG</sub> as input in addition to the region given by the LVC-Locator.

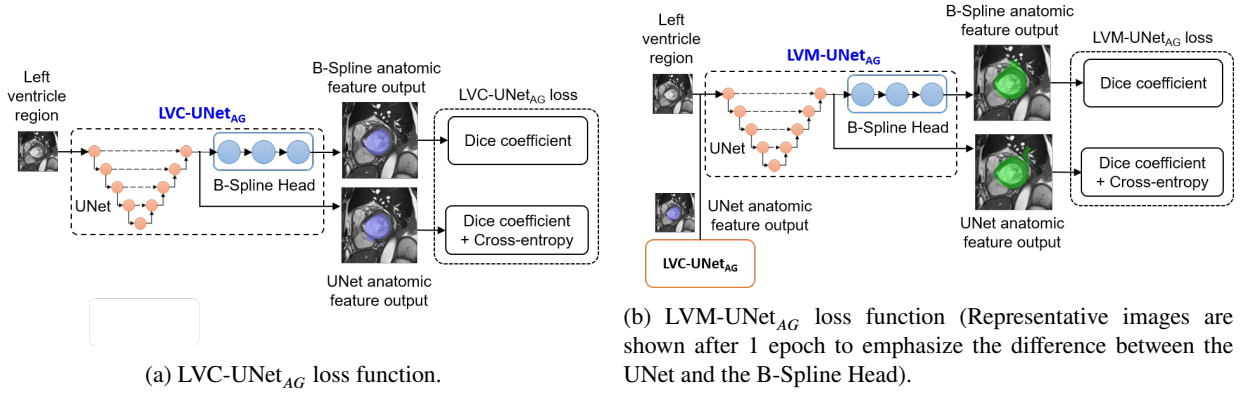
# Anatomically-guided deep learning for left ventricle geometry generation with uncertainty quantification based on short-axis MR images



(a) Anatomically accurate feature from U-Net with minimal impact from the B-spline head.

(b) Anatomically inaccurate feature from U-Net where the protrusion will be efficiently corrected by the B-spline head.

**Figure 3:** B-spline Head architecture details. LVM predictions are colored in green, UNet prediction contours are blue, B-spline curves are yellow, and control points are red. Results obtained after training our model for only one epoch.



(a) LVC-UNet<sub>AG</sub> loss function.

(b) LVM-UNet<sub>AG</sub> loss function (Representative images are shown after 1 epoch to emphasize the difference between the UNet and the B-Spline Head).

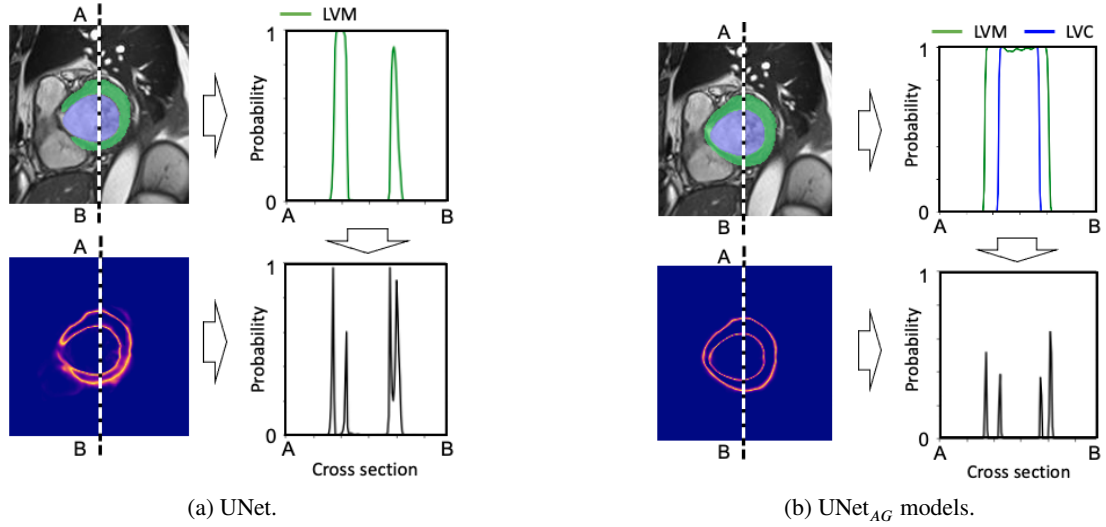
**Figure 4:** Loss function compositions for the LVC- and LVM-UNet<sub>AG</sub> models.

parameters are the ones associated with the internal UNet. Effectively, the B-spline head returns a smooth and constrained portion of the input image that has high probability of containing the anatomic feature of interest. Therefore, we propose using both the outputs of the internal UNet as well as the B-spline head during the training of these UNet<sub>AG</sub> models. As illustrated in Fig. 4, the loss function of either models is defined as:

$$\mathcal{L}_{\text{UNet}_{AG}} = -\frac{2}{|C|} \sum_{c \in C} \frac{\sum_i u_i^c l_i^c}{\sum_i u_i^c + \sum_i l_i^c} - \sum_{c \in C} \sum_i l_i^c \log(u_i^c) - \frac{2}{|C|} \sum_{c \in C} \frac{\sum_i s_i^c l_i^c}{\sum_i s_i^c + \sum_i l_i^c}, \quad (2)$$

where  $u$  is the output of the UNet,  $s$  is the output of the B-spline head,  $l$  is a one hot encoding of the label segmentation map,  $i$  is the voxel number in each map, and  $c \in C$  are the classes.  $u$ ,  $s$  and  $l$  are of size  $N_i \times N_c$ , where, as before,  $N_i$  and  $N_c$  are the number of voxels in each map and the number of classes, respectively. The classes used in Eq. (2) depend on the UNet<sub>AG</sub>: the LVC-UNet<sub>AG</sub> uses the LVC and background classes, while the LVM-UNet<sub>AG</sub> uses the LVM and background classes.

The composition of Dice coefficient and cross-entropy applied to the internal UNets works similarly to many other applications reported in the literature. In addition, we also use the Dice coefficient applied to the B-spline outputs while computing the overall loss and consequently while computing the gradients of the internal UNet. As we discussed, the B-spline head has no trainable parameters and its contribution to the loss is used to help the hyperparameter optimization of the internal UNet. This means that the B-spline head is an active regularizer for the training of the proposed anatomically-guided UNet models.



**Figure 5:** Uncertainty in geometry for UNet and UNet<sub>AG</sub> models.

### 3.3. Uncertainty in left ventricle geometry

Classic UNet models are trained to segment the input images into pre-defined regions. After a UNet model is trained, it will output probabilities for each region and the background on a pixel-by-pixel basis. Indeed, the anatomic features can be segmented based on an established probability threshold for each region. The top-left panel of Fig. 5a illustrates two regions of interest (LVM and LVC) identified by the UNet. The probabilities are very low for both LVM and LVC throughout the background (outside the region of interest), while the LVM probability increases significantly in the region within the inner and outer walls. In this case, the probability  $P_{\text{Wall-UNet}}$  for the inner and outer walls of the left ventricle can be obtained as:

$$P_{\text{Wall-UNet}}(x, y) = 1 - 2 \times |P_{\text{LVM}}(x, y) - 0.5|, \quad (3)$$

where  $x$  and  $y$  are the coordinates of each pixel; and  $P_{\text{LVM}}(\cdot)$  are the probability values assigned by the UNet to the “LVM” class. Similarly to the last layer in the UNet (softmax), Eq. (3) ensures that  $P_{\text{Wall-UNet}}$  is zero in the background and LVC and pixels with the highest probability can approach 1. The bottom panel of Fig. 5a shows these probabilities for the inner and outer walls. Inevitably, there is considerable overlap of the inner and outer wall probabilities. Unfortunately, this intersection will effectively reduce the ability to clearly distinguish the inner and outer wall.

A major advantage of the proposed approach is the quantification of uncertainty in the inner- and outer-walls of the left ventricle. We use the LVC-UNet<sub>AG</sub> model to obtain the probabilities of the inner wall and the LVM-UNet<sub>AG</sub> model to obtain the probabilities of the outer wall, as shown in the top panel of Fig. 5b, right. The LVM-UNet<sub>AG</sub> is informed by the LVC-UNet<sub>AG</sub>, and therefore the wall thickness estimate is significantly improved when compared to the classic UNet model. As the LVM-UNet<sub>AG</sub> output contains also the LVC probability, the left ventricle wall probability  $P_{\text{Wall-UNet}_{AG}}$  can be obtained as:

$$P_{\text{Wall-UNet}_{AG}}(x, y) = 1 - 2 \times |P_{\text{LVM-UNet}_{AG}}(x, y) - P_{\text{LVC-UNet}_{AG}}(x, y) - 0.5|, \quad (4)$$

where  $x$  and  $y$  are the coordinates of each pixel; and  $P_{\text{LVM-UNet}_{AG}}(\cdot)$  and  $P_{\text{LVC-UNet}_{AG}}(\cdot)$  are the probability values assigned by the LVM-UNet<sub>AG</sub> and LVC-UNet<sub>AG</sub> models, respectively. As before, the bottom panel of Fig. 5b, right shows the probabilities for the inner and outer walls. With the proposed approach, we observe that both the transition from background to left ventricle outer wall and the transition from inner wall to blood pool are very sharp.

## 4. Case-study

In this section we introduce the case study adopted to test the proposed anatomically-guided, geometry generation

pipeline. Details on the model training and results reproducibility are also presented to facilitate the adoption of our scheme by other researchers.

#### 4.1. Cardiac MRI data

In this work, we used the ACDC dataset [8, 9] to build our case study. Short-axis, cine MR images of 150 patients are provided in the ACDC dataset, including ground truth segmentations of LVM and LVC at end diastole (ED) and end systole (ES). As described in [9], the images were acquired in a 1.5 T (Siemens Aera) or a 3 T (Siemens Trio Tim) MRI scanner. In-plane pixel size varied from 1.34 mm  $\times$  1.34 mm to 1.68 mm  $\times$  1.68 mm, while slice thickness varied from 5 mm to 10 mm (in some cases including a 5 mm gap between slices).

Before training, validation, and testing are performed, the dataset is split into 100 training and 50 test patients. In order to overcome significant differences in contrast, the images' signal intensity is normalized between  $-1$  and  $1$  by: (1) scaling the central 98% of the pixel intensity distribution between  $-1$  and  $1$ ; and (2) truncating the tails of the intensity distribution at  $-1$  and  $1$ . In addition, all images are re-sampled to a uniform 1.25 mm  $\times$  1.25 mm in plane resolution using bi-cubic interpolation.

#### 4.2. Configuration of the UNet<sub>AG</sub> training

Training was carried out using a 5-fold cross-validation strategy. Accordingly, the data was divided into five groups, each containing the images related to twenty patients. Training is based on four groups while the remaining group is used for validation. Based on this strategy, five models were setup for each network (LVC-Locator, LVC-UNet<sub>AG</sub>, and LVM-UNet<sub>AG</sub>). Each training contained 300 epochs with a learning rate of  $5e-4 \cdot 0.985^{\text{epochs}}$ . In order to avoid overfitting and to increase the diversity of the dataset, we performed data augmentation using a sequence of rotations, scaling, horizontal/vertical mirror transformations, and elastic morphing [34] as outlined in [25]. Data augmentation was carried out offline and ten new images were created from each original image. In addition, to account for variability in the location of the image central point identified by the heart locator, a vertical and horizontal translation (randomly chosen from a uniform distribution between  $-10$  and  $10$  pixels) was added online in each epoch before the images were cropped.

Regarding model parameters and training time, the LVC-Locator, LVC-UNet<sub>AG</sub>, and LVM-UNet<sub>AG</sub> architectures have approximately 17.7M parameters each. Training of the LVC-Locator required  $\approx 67$  hours, whereas training of the UNet<sub>AG</sub> models required  $\approx 30$  hours each. Inference time for all the models is in the order of one second. All networks were trained using a Linux server configured with 64 Intel(R) Xeon(R) CPUs E5-2683 v4 at 2.10GHz, 132GB of RAM, 2 16GB NVIDIA Tesla P100 GPUs, and running Ubuntu 16.04.

#### 4.3. Results reproducibility

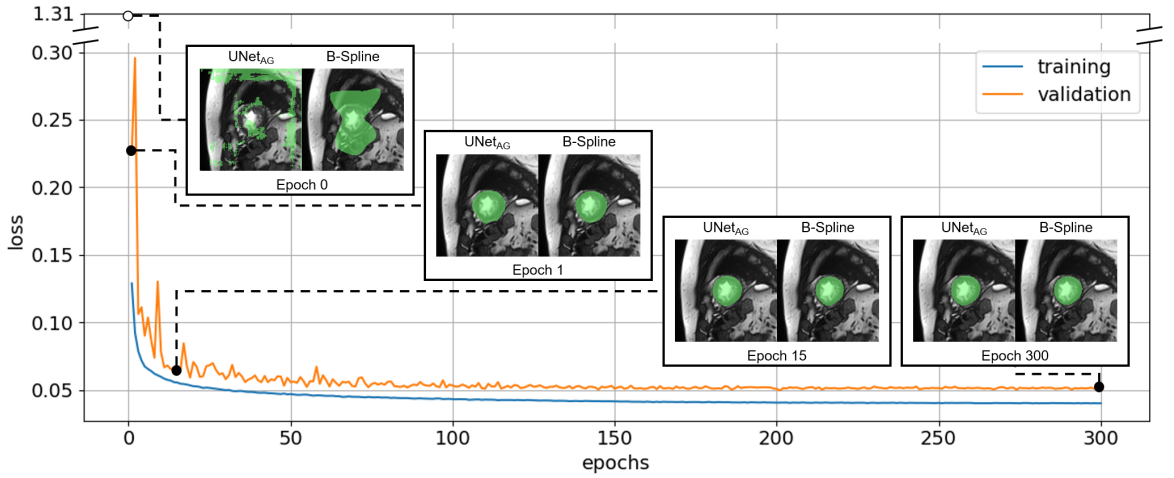
The proposed model is implemented in TensorFlow<sup>2</sup> (version 2.3.1) using the Python programming interface. The source codes and links to the data can be found in the following GitHub repository: [https://github.com/CBL-UCF/unet\\_ag](https://github.com/CBL-UCF/unet_ag).

### 5. Results and discussion

Four sample outputs of the UNet<sub>AG</sub> and the B-spline head at different epochs using the validation data are shown in Fig. 6, together with the evolution of the LVM-UNet<sub>AG</sub> training and validation losses across epochs. From epoch 0 to epoch 1, we notice the role of the proposed B-spline head in our UNet<sub>AG</sub> as it consolidates the matrix of probabilities generated by the UNet into a single region. As a result, it removes any islands and spikes that would otherwise be misclassified and improves the detected contours of the endocardial and epicardial surfaces. The constraint imposed by the B-spline head during training improves the hyperparameter optimization by guiding the UNet<sub>AG</sub> to focus on the region enclosed by the spline. As illustrated in Fig. 6, at epoch 0, when the UNet<sub>AG</sub> hyperparameters were simply randomly initialized, the B-spline head aggregates the prediction into a single fully-connected region, while the UNet<sub>AG</sub> output was noisy and affected by islands. After the first iteration, the model identifies the region of interest, and the subsequent epochs are used to refine the prediction. On top of generating smooth contours, this aggregation feature makes post-processing methods for islands removal unnecessary.

Our proposed method is focused on anatomically-guided geometry generation, as opposed to conventional segmentation. Nevertheless, Table 1 offers to the interested reader a quantitative comparison between our UNet<sub>AG</sub>, our

<sup>2</sup><https://www.tensorflow.org/>



**Figure 6:** LVM-UNet<sub>AG</sub> training and validation losses across epochs. Epoch 0 corresponds to the stage before training begins.

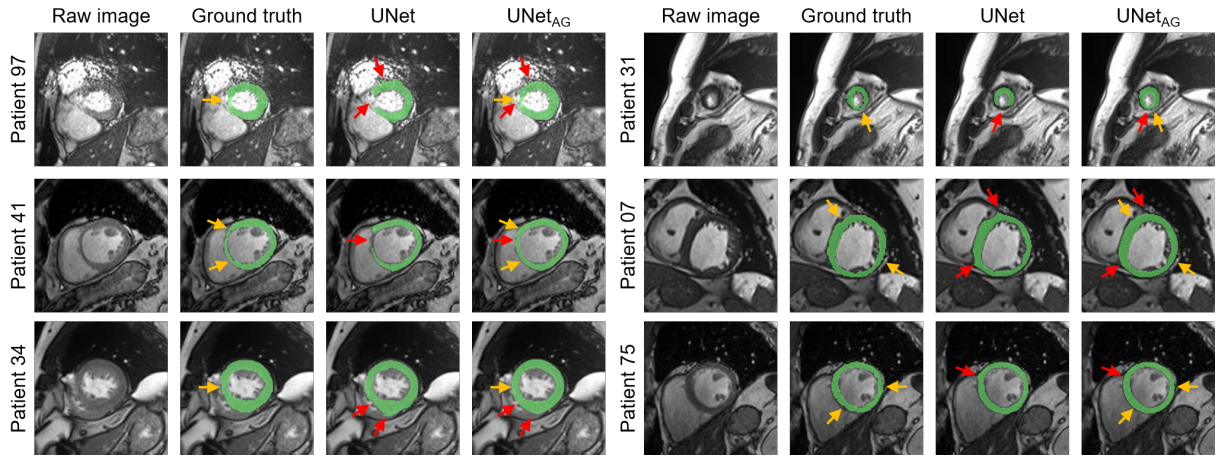
**Table 1**

Comparison among the proposed UNet<sub>AG</sub>, our previous work [27], and other representative frameworks reported in [9].

| Study                   | End Diastole |       |       |       | End Systole |       |       |       |
|-------------------------|--------------|-------|-------|-------|-------------|-------|-------|-------|
|                         | LVC          |       | LVM   |       | LVC         |       | LVM   |       |
|                         | DSC          | $d_H$ | DSC   | $d_H$ | DSC         | $d_H$ | DSC   | $d_H$ |
| Isensee et al. [25]     | 0.968        | 7.4   | 0.902 | 8.7   | 0.931       | 6.9   | 0.919 | 8.7   |
| Baumgartner et al. [16] | 0.963        | 6.5   | 0.892 | 8.7   | 0.911       | 9.2   | 0.901 | 10.6  |
| UNet <sub>AG</sub>      | 0.951        | 9.2   | 0.866 | 9.8   | 0.859       | 14.3  | 0.874 | 13.8  |
| UNet [27]               | 0.946        | 11.7  | 0.823 | 13.0  | 0.870       | 13.7  | 0.845 | 14.8  |
| Tziritas-Grinias [29]   | 0.948        | 8.9   | 0.794 | 12.6  | 0.865       | 11.6  | 0.801 | 14.8  |
| Yang et al. [15]        | 0.864        | 47.9  | N/A   | N/A   | 0.775       | 53.1  | N/A   | N/A   |

previous work [27], and four representative models that participated in the MICCAI 2017 challenge, as reported in [9] (additional comparisons with other approaches can be found in Table III reported in [9]). In the compared studies, Isensee et al. [25] presented a framework that averages 2D- and 3D-UNets predictions, Baumgartner et al. [16] utilized a 2D UNet, Tziritas and Grinias [29] evaluated an MRF model, and Yang et al. [15] assessed a 3D-FCN approach. This quantitative comparison is provided in terms of Dice similarity coefficient (DSC) and Hausdorff surface distance ( $d_H$ ). The reported metrics reflect only the degree of agreement with the adopted ground truth, but not the achievement of desired anatomical features. The comparison with a traditional approach (UNet [27]) shows the benefits of using the B-spline head during training. Overall the UNet<sub>AG</sub> improved the results obtained in our previous work [27], as evidenced by the decreased  $d_H$  and increased DSC in LVM for both cardiac phases and in LVC at end-diastole. This comparison reflects that our UNet<sub>AG</sub> model generates anatomically consistent geometries while achieving an agreement with the ground truth comparable to that of the best models presented during the MICCAI 2017 challenge.

Next, we discuss the ability of our UNet<sub>AG</sub> model to generate anatomically consistent geometries. In order to highlight the robustness of our approach, the results shown from this point forward were all obtained in a cross-validation fashion. In other words, the images used to generate the predictions were not present in the training set. Fig. 7 shows our model predictions alongside with the raw images, the ground truth, and a regular UNet prediction based on the model described in [27]. The yellow arrows in the ground truth columns indicate challenging regions in the ground truth images as, for example, image artifacts, jagged segmentation of the endocardium and epicardium, and uneven segmentations (these challenges are also noted in Fig. 1). The red arrows in the UNet columns highlight aspects of the predicted segmentations, such as lack of continuity of the ventricular wall and sharp corners, which are anatomically incorrect. Finally, the UNet<sub>AG</sub> column shows the geometries obtained with the newly proposed method. In order to assist with the evaluation of the improvements achieved with the current method, the yellow and red arrows

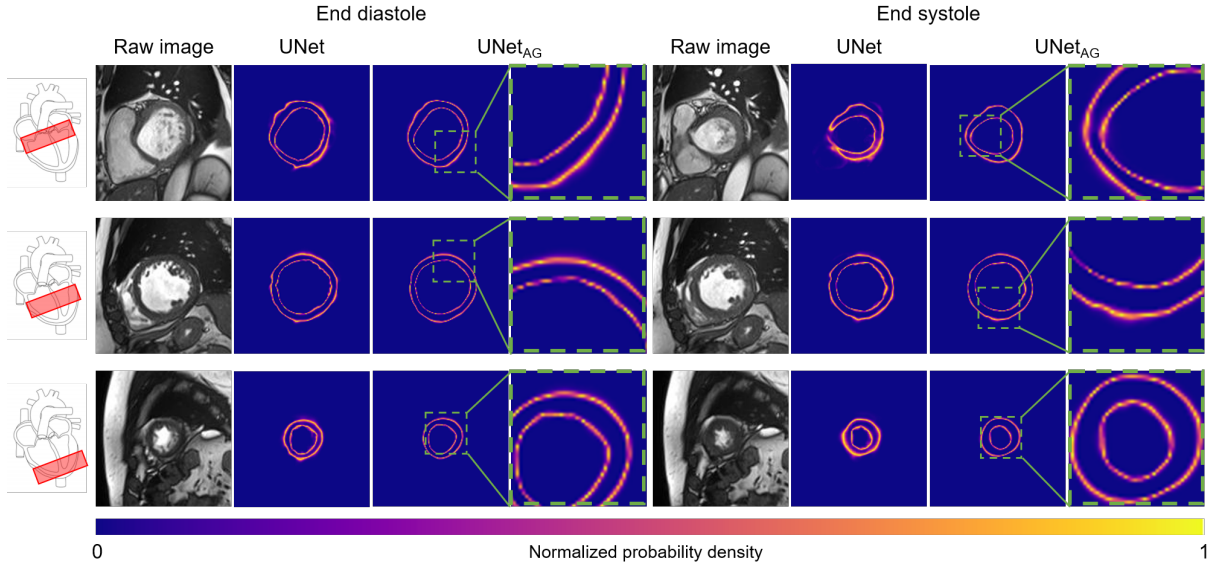


**Figure 7:** Raw image, ground truth, and segmentation results obtained with a regular UNet based on the model found in [27] and the proposed UNet<sub>AG</sub>. LVM predictions are noted in green. The yellow arrows indicate challenges in the ground truth labels and the red arrows point to problems found in the UNet predictions.

are reported at the same locations where the issues in the ground truth and UNet segmentations were detected. For example, in patients 97, 31, and 41, the UNet predictions underestimate the left ventricle myocardium and lead to unphysiological discontinuities in the left ventricle. These incorrect features are eliminated in the predictions generated by UNet<sub>AG</sub>. The B-spline curve that shapes the endocardial and epicardial contours enforce continuity in the predicted region. Furthermore, the use of the LVC-UNet<sub>AG</sub> output as an additional input channel for the LVM-UNet<sub>AG</sub> helps the proposed model to learn the anatomy constraints for the endocardial and the epicardial outlines. In other cases, the left ventricle myocardium can be overpredicted by the UNet model. For example, examining the segmentations for patients 97, 34, 07, and 75, it is possible to observe sharp corners and additional myocardium predicted by the UNet model, in particular close to the RV insertion points. In contrast, the proposed B-spline constraint prevents the UNet<sub>AG</sub> from overpredicting the left ventricle myocardium, restoring a physiologically accurate anatomy. The jaggedness present in the ground truth labels' contours is not anatomically correct and may be caused by limitations in the annotation tools and limited images' spatial resolution. For all cases shown in Fig. 7, the contours predicted by the UNet<sub>AG</sub> significantly reduce the unphysiological jaggedness leading to endocardial and epicardial outlines that are smoother than the ones generated by the regular UNet, and even smoother than the ground truth labeled contours. Instances highlighting this feature of the proposed UNet<sub>AG</sub> are indicated by the yellow arrows in the images for patients 41, 31, 07, and 75.

A major benefit of the proposed UNet<sub>AG</sub> is the improved uncertainty quantification of the geometry, measured as uncertainty in the location of the endocardial and epicardial contours. This uncertainty reflects the degree of difficulty in feature detection due, for example, to the low contrast between myocardium and background (apical slices), the presence of a highly corrugated endocardial trabecular structure (midventricular and apical slices), and/or the amount of artifacts (basal slices). Fig. 8 shows the comparison between the uncertainties quantified by a regular UNet versus the ones computed using the UNet<sub>AG</sub> at end-diastole and end-systole for basal (top row), mid-ventricular (middle row), and apical (bottom row) slices. The regions of high probability provided by the regular UNet are thicker than the ones provided by the UNet<sub>AG</sub> models. In addition, as illustrated in Fig. 8, the regular UNet may show discontinuities in the prediction in the basal slice at the end of systole. Regular UNets suffer from the fact that their outputs are probabilities for the different classes with no constraints that help distinguishing between classes. In contrast, in our approach, the LVM-UNet<sub>AG</sub> is informed by the LVC-UNet<sub>AG</sub> and their B-spline heads help the model to output smooth continuous contours. The refinement phase in the training helps reducing the regions of high uncertainty, which greatly improves the confidence in the generated results. As shown in Fig. 6, the majority of the computational effort is indeed dedicated to refining the prediction of the UNet<sub>AG</sub>.

Finally, as illustrated in Fig. 9, the uncertainty estimates provided by the proposed UNet<sub>AG</sub> models are also useful when analyzing the entire base-to-apex set of images at the end-diastolic and end-systolic cardiac phases. Fig. 9a presents a healthy subject, while Fig. 9b shows a patient affected by dilated cardiomyopathy. The smooth and sharply-defined regions of high probability help quantifying anatomical difference in health and disease and changes in wall



**Figure 8:** Raw image, UNet, and UNet<sub>AG</sub> normalized probability density of the predicted contours at end-diastole and end-systole for a basal (top), mid-ventricular (middle), and apical (bottom) slices. The UNet predictions are based on [27].

thickness during the cardiac cycle. Based on the predictions generated using the healthy subject data (Fig. 9a), it is possible to observe, as expected during cardiac contraction, a clear wall thickening from end diastole to end systole. Left ventricular wall thickening is not as pronounced in the predictions based on the data acquired in the patient affected by dilated cardiomyopathy (Fig. 9b) [35]. We also observe that the uncertainty in geometry increases (enlarged regions of high probability) in the predictions computed from the images of a patient affected by dilated cardiomyopathy compared to the predictions computed from a healthy subject data.

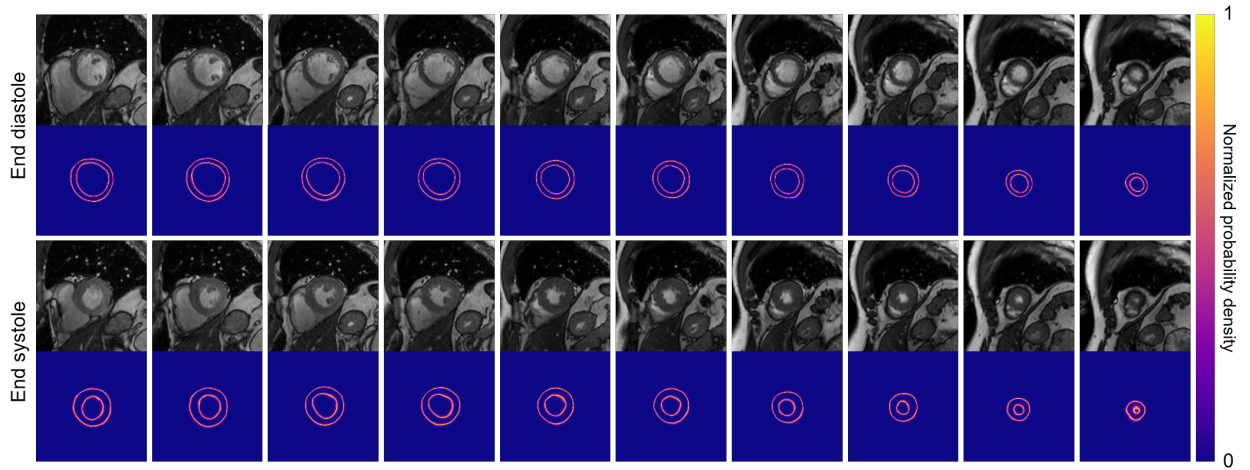
## 6. Summary and closing remarks

In this paper, we proposed an anatomically-guided deep learning approach for generating the geometry of the left ventricle based on short-axis MR images. In our implementation, we addressed challenges commonly found in predictions of left ventricular geometries such as discontinuity in anatomical features, excessive jaggedness and unevenness along myocardial walls, and poor uncertainty estimates.

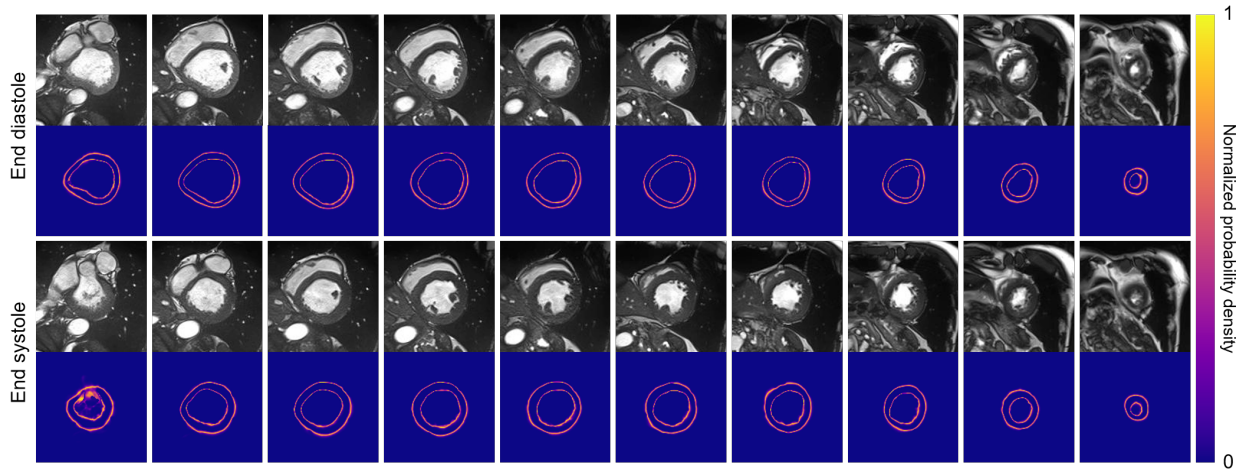
In order to address these challenges, we proposed a pipeline composed of three deep learning models, each with a specialized function. The first model is an LVC-Locator that receives raw MR images as input and focuses the geometry generation around the region of interest. Subsequently, two anatomically-guided deep learning models generate separately the geometry of the left ventricle cavity and the left ventricular myocardium. The anatomically-guided deep learning models are composed by a UNet followed by a B-spline head, which we called UNet<sub>AG</sub>. While the UNet is responsible for identifying the region of interest, the B-spline head works as a regularization layer during training. After training, the UNet<sub>AG</sub> models return the probability of each pixel belonging to one class (myocardium or cavity) with smooth and continuous probability contours. While the B-spline head has no trainable parameters, it contributes to the loss function used in the training of the underlying UNet. Although we use a UNet as our segmentation model, the proposed B-spline head can be further utilized with any neural network architecture, as it is an independent module.

Conventional UNet-based approaches are capable of segmenting MR images, but they may generate anatomically unrealistic features. In contrast, the proposed pipeline is capable of generating smooth, continuous, and accurate anatomical features. We demonstrate our framework capability using the ACDC dataset [8, 9], highlighting issues encountered with standard UNets and ground truth manual segmentations that are resolved by the UNet<sub>AG</sub> models. We observed that the introduction of the B-spline head in the UNet<sub>AG</sub> model has substantially improved the quality of the generated geometry while also accelerating training convergence.

The smoothness, continuity, anatomical consistency, and robust uncertainty quantification of the generated left ventricular geometries make our framework ideal for applications in patient-specific computational cardiology, where



(a) Patient 75 - Healthy



(b) Patient 05 - Affected by dilated cardiomyopathy

**Figure 9:** Base to apex (left to right) raw images and normalized probability density of the endocardial and epicardial contours predicted with the  $UNet_{AG}$  model at end-diastole and end-systole. A direct comparison between the contours computed for the healthy subject (top) and patient affected by dilated cardiomyopathy (bottom) reveals that the uncertainty is significantly lower when computed based on the healthy subject data.

anatomical fidelity and a measure of model uncertainty are critical.

The proposed pipeline also presents several limitations and opportunities for future improvement. Accuracy can be increased by optimizing the neural network architecture, e.g., the number of layers, number of neurons in each layer, and activation functions. This level of optimization was not pursued in the current work, but a neural architecture search [36–38] can be carried out to optimize the data-driven portions of the model. In this context, while currently there was no reason to optimize the number of B-spline head control points, this task can be explored in future applications. In terms of the application, we found that the performance of both conventional UNets and our proposed  $UNet_{AG}$  are still not optimal when segmenting the most basal and apical slices due to decreased image quality (e.g., lower contrast and increased artifacts) in these regions. These aspects need to be specifically targeted in future research.

We conclude by highlighting that, in the future, the proposed framework can be adapted and applied to generate the image-based geometry of other organs such as, for example, the liver, the kidney, and large blood vessels. To facilitate the dissemination and adoption of the presented methodology, the codes developed in this work has been made publicly available on GitHub ([https://github.com/CBL-UCF/unet\\_ag](https://github.com/CBL-UCF/unet_ag)) under the MIT License.

## Acknowledgements

This material is based upon work supported by the National Science Foundation under Award Number 2205043.

## CRedit authorship contribution statement

**Andre Von Zuben:** Methodology, Software, Formal Analysis, Investigation, Data Curation, Writing, Visualization. **Luigi E. Perotti:** Conceptualization, Methodology, Validation, Software, Formal Analysis, Investigation, Writing, Supervision, Funding Acquisition. **Felipe A. C. Viana:** Conceptualization, Methodology, Validation, Software, Formal Analysis, Investigation, Writing, Supervision, Funding Acquisition.

## References

- [1] S. A. Niederer, J. Lumens, N. A. Trayanova, Computational models in cardiology, *Nature Reviews Cardiology* 16 (2019) 100–111.
- [2] L. E. Perotti, I. A. Verzhbinsky, K. Moulin, T. E. Cork, M. Loecher, D. Balzani, D. B. Ennis, Estimating cardiomyofiber strain in vivo by solving a computational model, *Medical image analysis* 68 (2021) 101932.
- [3] H. Zou, S. Leng, C. Xi, X. Zhao, A. S. Koh, F. Gao, J. Le Tan, R.-S. Tan, J. C. Allen, L. C. Lee, et al., Three-dimensional biventricular strains in pulmonary arterial hypertension patients using hyperelastic warping, *Computer methods and programs in biomedicine* 189 (2020) 105345.
- [4] M. R. Pfaller, J. M. Hörmann, M. Weigl, A. Nagler, R. Chabiniok, C. Bertoglio, W. A. Wall, The importance of the pericardium for cardiac biomechanics: from physiology to computational modeling, *Biomechanics and modeling in mechanobiology* 18 (2019) 503–529.
- [5] V. Y. Wang, H. Lam, D. B. Ennis, B. R. Cowan, A. A. Young, M. P. Nash, Modelling passive diastolic mechanics with quantitative MRI of cardiac structure and function, *Medical image analysis* 13 (2009) 773–784.
- [6] F. Kong, S. C. Shadden, Automating Model Generation for Image-Based Cardiac Flow Simulation, *Journal of Biomechanical Engineering* 142 (2020). 111011. doi:10.1115/1.4048032.
- [7] F. Kong, N. Wilson, S. Shadden, A deep-learning approach for direct whole-heart mesh reconstruction, *Medical Image Analysis* 74 (2021) 102222. doi:https://doi.org/10.1016/j.media.2021.102222.
- [8] P.-M. Jodoin, A. Lalande, O. Bernard, O. Humbert, C. Zotti, F. Cervenansky, Automated cardiac diagnosis challenge, (accessed: 04.09.2020), 2017. URL: <https://acdc.creatis.insa-lyon.fr/>.
- [9] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P. A. Heng, I. Cetin, K. Lekadir, O. Camara, M. A. Gonzalez Ballester, G. Sanroma, S. Napel, S. Petersen, G. Tziritas, E. Grinias, M. Khened, V. A. Kollerathu, G. Krishnamurthi, M. M. Rohé, X. Pennec, M. Sermesant, F. Isensee, P. Jäger, K. H. Maier-Hein, P. M. Full, I. Wolf, S. Engelhardt, C. F. Baumgartner, L. M. Koch, J. M. Wolterink, I. Išgum, Y. Jang, Y. Hong, J. Patravali, S. Jain, O. Humbert, P. M. Jodoin, Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved?, *IEEE transactions on medical imaging* 37 (2018) 2514–2525.
- [10] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, C. I. Sánchez, A survey on deep learning in medical image analysis, *Medical Image Analysis* 42 (2017) 60–88. doi:https://doi.org/10.1016/j.media.2017.07.005.
- [11] C. Chen, C. Qin, H. Qiu, G. Tarroni, J. Duan, W. Bai, D. Rueckert, Deep Learning for Cardiac Image Segmentation: A Review, *Frontiers in Cardiovascular Medicine* 7 (2020) 25. doi:10.3389/fcvm.2020.00025.
- [12] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W. M. Wells, A. F. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Springer International Publishing, Cham, 2015, pp. 234–241.
- [13] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, J. Liang, Unet++: A nested U-Net architecture for medical image segmentation, in: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, 2018, pp. 3–11.
- [14] S. Li, Y. Chen, S. Yang, W. Luo, Cascade Dense-Unet for Prostate Segmentation in MR Images, in: D.-S. Huang, V. Bevilacqua, P. Premaratne (Eds.), *Intelligent Computing Theories and Application*, Springer International Publishing, Cham, 2019, pp. 481–490.
- [15] X. Yang, C. Bian, L. Yu, D. Ni, P.-A. Heng, Class-balanced deep neural network for automatic ventricular structure segmentation, in: M. Pop, M. Sermesant, P.-M. Jodoin, A. Lalande, X. Zhuang, G. Yang, A. Young, O. Bernard (Eds.), *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*, Springer International Publishing, Cham, 2018, pp. 152–160.
- [16] C. F. Baumgartner, L. M. Koch, M. Pollefeys, E. Konukoglu, An exploration of 2D and 3D deep learning techniques for cardiac MR image segmentation, in: M. Pop, M. Sermesant, P.-M. Jodoin, A. Lalande, X. Zhuang, G. Yang, A. Young, O. Bernard (Eds.), *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*, Springer International Publishing, Cham, 2018, pp. 111–119.
- [17] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, D. Rueckert, Attention U-Net: Learning where to look for the pancreas, 2018. [arXiv:1804.03999](https://arxiv.org/abs/1804.03999).
- [18] S. Jetley, N. A. Lord, N. Lee, P. Torr, Learn to pay attention, in: *International Conference on Learning Representations*, 2018. URL: <https://openreview.net/forum?id=HyzbhfWRW>.
- [19] S. Li, J. Zhang, C. Ruan, Y. Zhang, Multi-Stage Attention-Unet for Wireless Capsule Endoscopy Image Bleeding Area Segmentation, in: *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2019, pp. 818–825. doi:10.1109/BIBM47256.2019.8983292.
- [20] M. Islam, V. S. Vibashan, V. J. M. Jose, N. Wijethilake, U. Utkarsh, H. Ren, Brain Tumor Segmentation and Survival Prediction Using 3D Attention UNet, in: A. Crimi, S. Bakas (Eds.), *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, Springer International Publishing, Cham, 2020, pp. 262–272.
- [21] H. Cheng, J. Lu, M. Luo, W. Liu, K. Zhang, PTANet: Triple Attention Network for point cloud semantic segmentation, *Engineering Applications of Artificial Intelligence* 102 (2021) 104239. doi:https://doi.org/10.1016/j.engappai.2021.104239.

- [22] E. Ferdian, A. Suinesiaputra, K. Fung, N. Aung, E. Lukaschuk, A. Barutcu, E. Maclean, J. Paiva, S. K. Piechnik, S. Neubauer, E. P. Steffen, A. Y. Alistair, Fully Automated Myocardial Strain Estimation from Cardiovascular MRI-tagged Images Using a Deep Learning Framework in the UK Biobank, *Radiology: Cardiothoracic Imaging* 2 (2020) e190032. doi:10.1148/ryct.2020190032.
- [23] M. Khened, V. A. Kollerathu, G. Krishnamurthi, Fully convolutional multi-scale residual DenseNets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers, *Medical Image Analysis* 51 (2019) 21–45. doi:https://doi.org/10.1016/j.media.2018.10.004.
- [24] J. Zhao, M. Dang, Z. Chen, L. Wan, DSU-Net: Distraction-Sensitive U-Net for 3D lung tumor segmentation, *Engineering Applications of Artificial Intelligence* 109 (2022) 104649. doi:https://doi.org/10.1016/j.engappai.2021.104649.
- [25] F. Isensee, P. F. Jaeger, P. M. Full, I. Wolf, S. Engelhardt, K. H. Maier-Hein, Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features, in: M. Pop, M. Sermesant, P.-M. Jodoin, A. Lalande, X. Zhuang, G. Yang, A. Young, O. Bernard (Eds.), *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*, Springer International Publishing, Cham, 2018, pp. 120–129.
- [26] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, K. H. Maier-Hein, nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation, *Nature methods* (2020). doi:10.1038/s41592-020-01008-z.
- [27] A. V. Zuben, K. Heckman, F. A. C. Viana, L. E. Perotti, A multi-step machine learning approach for short axis MR images segmentation, in: *11th Biennial Meeting on Functional Imaging and Modeling of the Heart*, Springer, Virtual Event, 2021. doi:10.1007/978-3-030-78710-3\_13.
- [28] D. M. Popescu, H. G. Abramson, R. Yu, C. Lai, J. K. Shade, K. C. Wu, M. Maggioni, N. A. Trayanova, Anatomically informed deep learning on contrast-enhanced cardiac magnetic resonance imaging for scar segmentation and clinical feature extraction, *Cardiovascular Digital Health Journal* (2021). doi:https://doi.org/10.1016/j.cvdhj.2021.11.007.
- [29] E. Grinias, G. Tziritas, Fast fully-automatic cardiac segmentation in MRI using MRF model optimization, substructures tracking and B-Spline smoothing, in: M. Pop, M. Sermesant, P.-M. Jodoin, A. Lalande, X. Zhuang, G. Yang, A. Young, O. Bernard (Eds.), *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*, Springer International Publishing, Cham, 2018, pp. 91–100.
- [30] M. Beetz, A. Banerjee, V. Grau, Biventricular surface reconstruction from cine MRI contours using point completion networks, in: *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, 2021, pp. 105–109. doi:10.1109/ISBI48211.2021.9434040.
- [31] A. Banerjee, J. Camps, E. Zacur, C. M. Andrews, Y. Rudy, R. P. Choudhury, B. Rodriguez, V. Grau, A completely automated pipeline for 3D reconstruction of human heart from 2D cine magnetic resonance slices, *Philos Trans A Math Phys Eng Sci* 379 (2021) 20200257.
- [32] F. Kong, N. Wilson, S. Shadden, A deep-learning approach for direct whole-heart mesh reconstruction, *Medical Image Analysis* 74 (2021) 102222. doi:https://doi.org/10.1016/j.media.2021.102222.
- [33] A. Davies, P. Samuels, *An introduction to computational geometry for curves and surfaces*, Oxford University Press, Inc., 1996.
- [34] P. Y. Simard, D. Steinkraus, J. C. Platt, Best practices for convolutional neural networks applied to visual document analysis., in: *Icdar*, volume 3, 2003.
- [35] Y. Yu, S. Yu, X. Tang, H. Ren, S. Li, Q. Zou, F. Xiong, T. Zheng, L. Gong, Evaluation of left ventricular strain in patients with dilated cardiomyopathy, *Journal of International Medical Research* 45 (2017) 2092–2100.
- [36] K. Kandasamy, W. Neiswanger, J. Schneider, B. Poczos, E. P. Xing, Neural architecture search with Bayesian optimisation and optimal transport, in: S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett (Eds.), *Advances in Neural Information Processing Systems* 31, Curran Associates, Inc., 2018, pp. 2016–2025.
- [37] C. Liu, B. Zoph, M. Neumann, J. Shlens, W. Hua, L.-J. Li, L. Fei-Fei, A. Yuille, J. Huang, K. Murphy, Progressive neural architecture search, in: *The European Conference on Computer Vision (ECCV)*, Computer Vision Foundation, Munich, Germany, 2018.
- [38] T. Elsken, J. H. Metzen, F. Hutter, Neural architecture search: a survey, *Journal of Machine Learning Research* 20 (2019) 1–21.