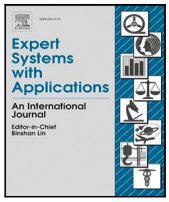
Expert Systems With Applications 257 (2024) 125005

Contents lists available at ScienceDirect

Expert Systems With Applications

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



HOCM-Net: 3D coarse-to-fine structural prior fusion based segmentation network for the surgical planning of hypertrophic obstructive cardiomyopathy

Jiawei Zhang [a,b,c], Xiaodong Wang [g,e], Hailong Qiu [b], Yanchun Zhang [c,a], Weihong Han [a], Jialin Wang [f], Tianchen Wang [d], Yiyu Shi [d], Meiping Huang [b], Jian Zhuang [b], Huiming Guo [b], Xiaowei Xu [b,*]

[a] The Department of New Networks, Pengcheng Laboratory, Shenzhen, China
[b] Guangdong Cardiovascular Institute, Guangdong Provincial Key Laboratory of South China Structural Heart Disease, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences) Southern Medical University, Guangzhou, 510080, China
[c] School of Computer Science and Technology, Zhejiang Normal University, Jinhua, China
[d] Department of Computer Science and Engineering, University of Notre Dame, IN, United States
[e] School of Computer Science and Technology, Xidian University, Xi'an, China
[f] Institute of Brain-Inspired Circuits and Systems (iBiCAS), Fudan University, Shanghai, China
[g] Cancer Center, Jinshan Hospital, Fudan University, Shanghai, China

ARTICLE INFO

Keywords:
Hypertrophic cardiomyopathy
Medical image segmentation
Surgical planning
Structural prior feature
Coarse-to-fine

ABSTRACT

Hypertrophic obstructive cardiomyopathy (HOCM) is a leading cause of sudden cardiac death in young people, and septal myectomy surgery has been recognized as the gold standard for non-pharmacological therapy of HOCM. Recently, 3D printing has been widely adopted to assist in the intuitive surgical planning of septal myectomy surgery. However, the process requires experienced surgeons to manually segment the excised myocardium (EM), which is time-consuming and expensive. Therefore, there is a strong demand for automated EM segmentation. This task poses a challenge due to the limited distinction between the EM and the remaining myocardium (RM), since both are components of the myocardium (Myo) without discrimination in voxel intensity and the boundaries of EM are determined by the positional relationships of certain cardiac sub-structures. Furthermore, the lack of publicly available datasets hampers research progress in this area. In this paper, we propose a 3D coarse-to-fine structural prior fusion based network named HOCM-Net for EM segmentation in the surgical planning of HOCM. HOCM-Net is a two-stage approach consisting of a coarse stage for localization of the heart and a fine stage for refined EM segmentation. Meanwhile, we introduce two modules to improve the segmentation, a 3D coarse-to-fine feature fusion module and a 3D structural prior feature fusion. The first one, inspired from the clinical practice, combines the structural prior feature of left ventricle (LV), right ventricle (RV), and Myo segmentation result in the coarse stage to refine the EM segmentation in the fine stage. The second one, for the purpose of coarse-to-fine feature fusion, employs two auxiliary scale features (one feature at a higher scale, one feature at a lower scale) to enhance the current scale feature at each layer and thus to improve the overall segmentation. To evaluate the performance of HOCM-Net, we collected a dataset comprising 40 CT volumes from HOCM patients who underwent septal myectomy surgery. Experimental results demonstrate that HOCM-Net surpasses state-of-the-art methods and exhibits promising potential for clinical applications. Additionally, we have made the dataset and the code publicly available to facilitate further research on this fascinating and challenging problem.

1. Introduction

Hypertrophic cardiomyopathy is a prevalent inherited cardiovascular disease, affecting approximately one in 500 individuals in the general population (Maron, 2018). It is the leading cause of sudden cardiac death among young people, primarily due to malignant arrhythmias (Maron & Maron, 2013). Hypertrophic cardiomyopathy is

* Corresponding author.
E-mail addresses: yanchun.zhang@vu.edu.au (Y. Zhang), hanweihong@gzhu.edu.cn (W. Han), guohuiming@gdph.org.cn (H. Guo), xiao.wei.xu@foxmail.com (X. Xu).

https://doi.org/10.1016/j.eswa.2024.125005
Received 3 July 2023; Received in revised form 13 July 2024; Accepted 4 August 2024
Available online 8 August 2024

characterized by asymmetric left ventricular hypertrophy in the absence of other underlying causes, typically identified by a maximal end-diastolic left ventricle (LV) wall thickness of 15 mm or more. Approximately 70% of hypertrophic cardiomyopathy patients experience left ventricular outflow obstruction at rest or under provocation, known as hypertrophic obstructive cardiomyopathy (HOCM) (Rowin, Maron, Olivotto, & Maron, 2017). For HOCM patients who have drug-refractory symptoms, the primary treatment method to relieve left ventricular outflow tract obstruction is septal myectomy surgery. This surgical procedure involves removing a small portion of the thickened septal wall. However, due to the intricate geometry of the heart and the highly diverse phenotypic expression of HOCM, septal myectomy surgery presents significant challenges that demand considerable experience and expertise. For instance, studies have demonstrated that lower surgeon volume is associated with worse outcomes, including longer hospital stays and higher mortality rates (Kim et al., 2016). Therefore, surgical planning is typically performed to address this issue, involving a meticulous examination of the patient's CT or MRI volumes in a slice-by-slice manner. By doing so, surgeons can gain insight into the patient's cardiac structure, enabling them to make a rough estimation of the amount of excised myocardium (EM). However, this method is still not entirely intuitive or accurate, requiring substantial effort to precisely estimate EM during the actual surgery.

Recently, the utilization of 3D printing in assisting septal myectomy surgery planning has become increasingly prevalent (Bartel, Rivard, Jimenez, Mestres, & Müller, 2018; Farooqi & Mahmood, 2018; Hamatani et al., 2017; Henn & Mokadam, 2021; Ma et al., 2021; Veselka, Adla, Adlova, & Bartel, 2018). Fig. 2 provides an overview of the processing pipeline. Initially, the original CT volumes undergo manual segmentation performed by surgeons. Subsequently, the surgeons can easily obtain the position relationship between EM and the myocardium (or the remaining myocardium (RM)), and assess the 3D size of EM. Ultimately, the surgeon can directly simulate the surgery using the printed model. It is important to note that the printed model can accurately replicate the surgical field and EM, as illustrated in the final row. Nevertheless, this pipeline necessitates highly skilled surgeons to conduct manual segmentation, which is time-consuming and costly, thereby impeding its further advancement and widespread adoption in clinical practice. Consequently, there is a strong demand for automated EM segmentation in the current clinical setting.

However, there are challenges and complexities associated with automatic EM segmentation, as we have observed through our experience with manual segmentation in clinical practice (Ma et al., 2021). EM segmentation specifically requires the precise separation of EM from RM, implying that the segmentation task involves distinguishing between EM and RM. However, there is no inherent visual distinction between EM and RM, as both regions belong to the myocardium (Myo) and exhibit similar voxel intensity values. In clinical practice, surgeons must consider the spatial relationship between important cardiac substructures such as the left ventricle (LV), right ventricle (RV), and Myo in order to accurately define boundaries. Fig. 1 illustrates a CT volume example and the corresponding ground truth, where the region containing EM is enclosed within red rectangle boxes, and it can be observed that the voxel intensity of EM is nearly no local discrimination from that of RM.

On the other hand, there is currently a lack of related research and publicly available datasets for automatic EM segmentation. Existing studies related to hypertrophic obstructive cardiomyopathy (HOCM) (Hodges et al., 2019; Ingles et al., 2019; Zhou et al., 2021) mainly focus on HOCM diagnosis, the relationship between HOCM and genetic factors, and surgical advancements in treating hypertrophic cardiomyopathy. Some studies (Ma et al., 2021) have started exploring the application of 3D printing to enhance the surgical process for HOCM, but these 3D printed models are manually annotated and proprietary. Other tasks (Bernard et al., 2018; Zhuang & Shen, 2016) related to heart segmentation primarily target whole heart segmentation or the segmentation of the left ventricle (LV). Despite the inclusion of MRI images from HOCM patients in the ACDC dataset (Bernard et al., 2018), it still lacks manual annotations for EM. The scarcity of public datasets can be attributed to the time-consuming and extensive manual segmentation required by experienced surgeons, which hampers further research on this intriguing topic within the scientific community.

In this paper, we present HOCM-Net, a coarse-to-fine structural prior fusion based EM segmentation network for surgical planning in cases of HOCM. To address the challenge of no local discrimination, we introduce a 3D structural prior feature fusion module, which incorporates prior medical knowledge to enhance EM segmentation. Specifically, we fuse the structural prior information (LV, RV, and Myo segmentation) with the input image to facilitate the identification of the boundary between EM and RM. This optimization is motivated by the fact that experienced surgeons typically consider LV, RV, and the entire Myo during manual segmentation of EM and RM. Additionally, we introduce a 3D coarse-to-fine feature fusion module, which employs two auxiliary scale features (one feature at a higher scale, one feature at a lower scale) to enhance the current scale feature at each layer and thus to improve the overall segmentation. Furthermore, we have curated a HOCM segmentation dataset consisting of 40 CT volumes from HOCM patients who underwent septal myectomy surgery at Guangdong Provincial People's Hospital between January 6, 2020, and February 5, 2021. This dataset has been made publicly available to foster further research in the field. To the best of our knowledge, this is the first publicly accessible dataset that includes precise manual labels for EM, specifically tailored for surgical planning in cases of HOCM.

Our contributions can be summarized as follows:

- We introduce HOCM-Net, a fusion-based EM segmentation network for surgical planning in HOCM (Hypertrophic Obstructive Cardiomyopathy). HOCM-Net utilizes both the 3D coarse-to-fine feature fusion module and the 3D structural prior feature fusion module to enhance segmentation performance. In addition, we have made our code publicly available[1] to facilitate future research.
- We introduce a 3D structural prior feature fusion module, which combines the structural information of left ventricle (LV), right ventricle (RV), and Myo segmentation result in the coarse stage to refine the EM segmentation in the fine stage.
- We introduce a 3D coarse-to-fine feature fusion module, which employs two auxiliary scale features (one feature at a higher scale, one feature at a lower scale) to enhance the current scale feature at each layer and thus to improve the overall segmentation.
- We have conducted extensive experiments, including quantitative evaluations and assessments by surgeons. The experimental results demonstrate that our method outperforms current state-of-the-art techniques and holds great potential for clinical applications in HOCM surgical planning.
- Additionally, we have curated and released the first publicly available dataset[2] for EM segmentation for surgical planning in cases of HOCM (HOCM24) to foster research within the community. This dataset comprises 40 CT volumes from HOCM patients who have undergone septal myectomy surgery.

The remainder of the paper is organized as follows. Section 2 describes the related works. Section 3 demonstrates the proposed cardiac CT image dataset, HOCM24. Section 4 introduces the architecture of HOCM-Net. Section 5 presents the experiments, results, and discussion, followed by the conclusion Section 6 and acknowledgments.

[1] https://github.com/JerRuy/HOCM-Net.

[2] https://github.com/XiaoweiXu/HOCM-Excised-Myocardium-Segmentation.

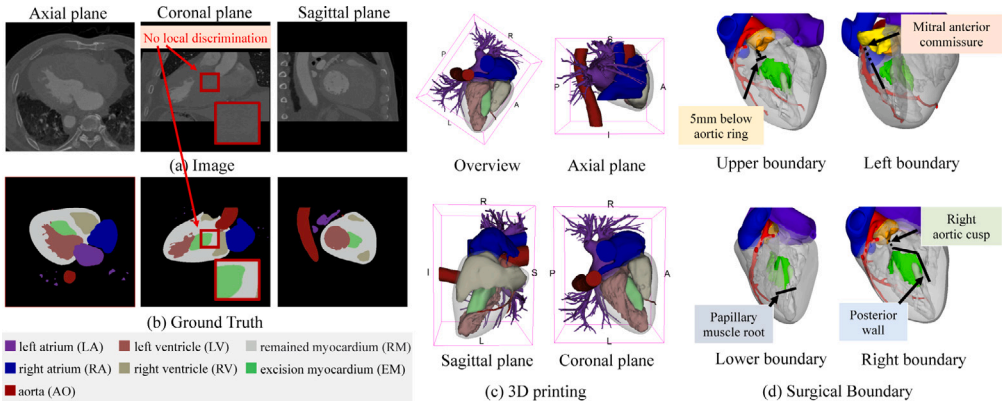


Fig. 1. Example CT volumes and ground truth of our proposed HOCM segmentation dataset: (a) three views of CT volumes, (b) the corresponding ground truth of (a), (c) three 3D views of the 3D segmentation results, and (d) 3D visualization of the surgical boundaries of excised myocardium (EM). Besides EM and remaining myocardium (RM), we also labeled five substructures of the heart including left atrium (LA), left ventricle (LV), right atrium (RA), right ventricle (RV), and aorta (Ao) to promote potential research on HOCM.

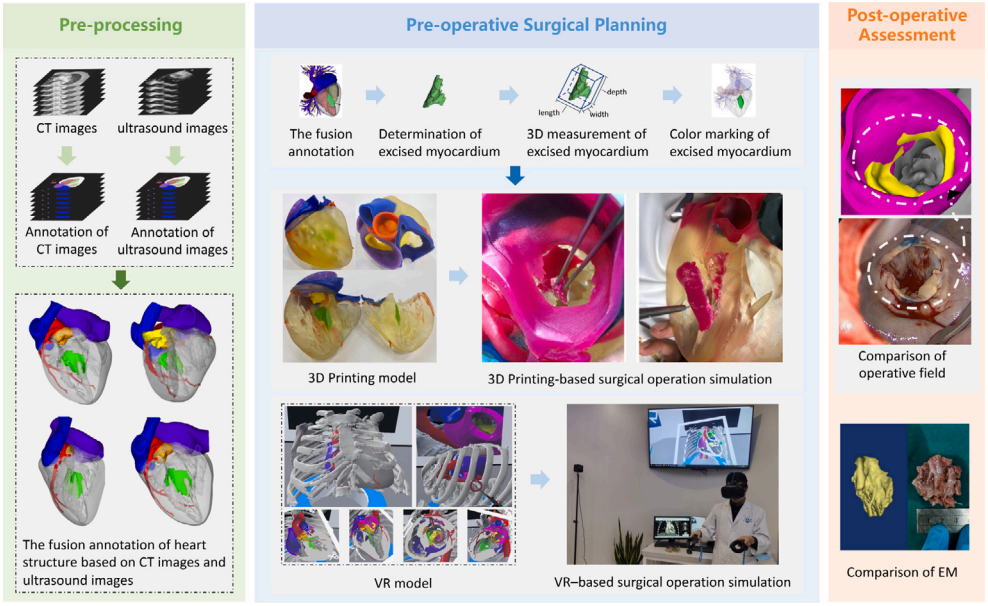


Fig. 2. Schematic illustration of the 3D printing-based and VR-based surgical planning of patient with HOCM (Ma et al., 2021). The overall process contains three steps: Pre-processing, 3D printing–based and VR–based pre-operative surgical planning, and Post-operative assessment.

2. Related work

Hypertrophic cardiomyopathy is a complex and heterogeneous genetic heart disease characterized by unexplained left ventricular hypertrophy. Over the past decade, significant progress has been made in the diagnosis and management of this condition (Maron & Maron, 2013; Mosqueira, Smith, Bhagwan, & Denning, 2019; Pičulin et al., 2022). Hypertrophic obstructive cardiomyopathy (HOCM) is a subtype of hypertrophic cardiomyopathy characterized by dynamic left ventricular outflow tract (LVOT) obstruction. In recent years, significant progress has been made in the diagnosis and treatment of this challenging cardiac condition (Iacovoni et al., 2012; Maron et al., 2003). In the realm of diagnosis, the use of advanced imaging modalities, such as echocardiography (Tuohy, Kaul, Song, Nazer, & Heitner, 2020) and cardiovascular magnetic resonance (CMR) imaging (Kamp et al., 2021), has become increasingly valuable. Echocardiography, particularly with the incorporation of Doppler techniques, allows for the detailed assessment of LVOT gradients (Geske, Sorajja, Ommen, & Nishimura, 2009; Pollick, Shmueli, Maalouf, & Zadikany, 2020), the extent of myocardial hypertrophy (Cui et al., 2021), and the degree of systolic anterior motion of the mitral valve (Vainrib et al., 2021). CMR, on the other hand, provides comprehensive structural and functional information, enabling the identification of high-risk features, such as myocardial fibrosis, that may guide risk stratification and treatment decisions. The diagnosis and management of hypertrophic obstructive cardiomyopathy (HOCM) have witnessed significant advancements in recent years, with the integration of machine learning (ML) and deep learning (DL) techniques playing a pivotal role. In the realm of diagnosis, ML and DL methods have been leveraged to enhance the interpretation of echocardiographic (Tuohy et al., 2020) and cardiovascular magnetic resonance (CMR) imaging data (Kamp et al., 2021). Automated algorithms have demonstrated the ability to accurately quantify left ventricular outflow tract (LVOT) gradients, characterize the degree of myocardial hypertrophy, and detect the presence of systolic anterior motion of the mitral valve — all critical parameters in the evaluation of HOCM. Furthermore, DL methods have been developed to identify high-risk features (Barker et al., 2022; Pičulin et al., 2022), such as myocardial fibrosis, which can be challenging to detect using conventional imaging techniques. By integrating these advanced image analysis capabilities, clinicians can now make more informed decisions regarding risk stratification and treatment planning. The application of ML and DL methods

Table 1
The difference analysis between our proposed cardiac image dataset (HOCM24) and existing datasets (SCD, RVSC, MM-WHS, and LVQuan19). Existing datasets tend to provide segmentation annotations for common substructures of the heart. Besides above, our dataset further provides annotations for the excised myocardium for the septal myectomy surgery of patients with HOCM. It marks the target of the heart segmentation, which has gradually moved from understanding the heart structure to clinical application.

Dataset	Year	Modalities	Number of images	Segmentation targets
SCD (Radau et al., 2009)	2009	MRI	45	Left ventricle
RVSC (Petitjean et al., 2015)	2012	MRI	48	Right ventricle
MM-WHS (Zhuang & Shen, 2016)	2017	CT/MRI	200	Whole heart
LVQuan19 (Xue, Brahm, Pandey, Leung, & Li, 2018)	2019	MRI	56	Left ventricle
HOCM24 (Ours)	2024	CT	40	Specific lesions*

The specific lesions here refer to the excised myocardium for the septal myectomy surgery of patients with HOCM.

have also extended to the genetic evaluation of HOCM (Khurshid et al., 2023; Maron et al., 2022). Automated algorithms have been trained to analyze genetic sequencing data, enabling the rapid and accurate identification of disease-causing mutations. This has not only improved the diagnostic process but also facilitated cascade screening of family members, a crucial step in the early detection and management of this inheritable condition. In terms of treatment, ML and DL methods have been employed to optimize patient selection and guide therapeutic decision-making (Augusto et al., 2021). For example, predictive models have been developed to identify patients most likely to benefit from septal reduction therapies (Ma et al., 2021), such as surgical myectomy or alcohol septal ablation. By incorporating a wide range of clinical, imaging, and genetic data, these models can provide personalized risk-benefit assessments, ultimately improving procedural outcomes and patient-centered care. Moreover, the use of DL methods in the interpretation of post-procedural data (Ouyang et al., 2024), such as echocardiographic and hemodynamic measurements, has enabled more accurate assessment of treatment efficacy and the identification of potential complications or suboptimal responses. Looking to the future, the continued integration of ML and DL techniques in the management of HOCM is expected to lead to further advancements. Ongoing research is exploring the development of virtual reality simulations (Jung et al., 2022), personalized computational models (Mancio et al., 2022; O'hara et al., 2022), and closed-loop optimization algorithms (Ma et al., 2021) to enhance surgical planning, device selection, and medication management. To the best of our knowledge, we are the first to develop a HOCM segmentation methods for the surgical planning of the surgery of the patient with HOCM.

In recent years, there has been significant progress in the field of cardiac image analysis, with the emergence of diverse cardiac image datasets encompassing different modalities such as CT, MRI, and ultrasound. As shown in Table 1, we compares our HOCM Segmentation Dataset (HOCM24) with existing image dataset (SCD, RVSC, MM-WHS, and LVQuan19) for heart segmentation. Existing datasets tend to provide segmentation annotations for common sub-structures of the heart. For instance, the MICCAI'17 multi-modality whole heart segmentation (MM-WHS) challenge (Zhuang et al., 2019) has played a crucial role in facilitating research advancements. This challenge provided 120 multi-modality cardiac images acquired in real clinical settings, creating an open and fair competition platform for research groups to evaluate and validate their methods. SCD (Sunnybrook Cardiac Data) (Radau et al., 2009): The SCD dataset, published in 2009, is a cardiac MRI dataset that provides segmentation annotations for the left ventricle. It contains 45 MRI images. RVSC (Right Ventricular Segmentation Challenge) (Petitjean et al., 2015): The RVSC dataset, published in 2012, is focused on the segmentation of the right ventricle from cardiac MRI images. It contains 48 MRI scans. LVQuan19 (Left Ventricle Quantification Challenge) (Xue et al., 2018): LVQuan19, published in 2019, is a cardiac MRI dataset that specializes in the segmentation of the left ventricle. It contains 56 MRI scans. Besides above commonly-seen cardiac sub-structures, our dataset further provides annotations for the resected septum for the septal myectomy surgery of patients with HOCM. It marks the target of the heart segmentation, which has gradually moved from understanding the heart sub-structures to clinical application. However, the lack of publicly available HOCM segmentation datasets with manual annotations of EM poses a challenge for the further validation of EM segmentation techniques. The labeling of HOCM requires extensive experience in septal myectomy surgery. To the best of our knowledge, we are the first to develop a HOCM segmentation dataset that includes cardiac images and corresponding EM labels.

Recently, a multitude of deep learning-based models (Çiçek, Abdulkadir, Lienkamp, Brox, & Ronneberger, 2016; Menegotto, Becker, & Cazella, 2021; Pandey et al., 2022; Ronneberger, Fischer, Brox, Brox, & Brox, 2015; Zhang, Jin, Xu, Xu, & Zhang, 2018; Zhang, Zhang, & Xu, 2021) have shown great potential in medical image analysis. For example, UNEt TRansformers (UNETR) (Hatamizadeh et al., 2021, 2022) utilizes a transformer as the encoder to learn sequence representations of the input volume and effectively capture the global multi-scale information, which achieves promising result on various biomedical image segmentation tasks. In particular, the nnU-Net (no new U-Net) (Isensee, Jaeger, Kohl, Petersen, & Maier-Hein, 2021) adjusts all hyperparameters based on the characteristics of a given dataset without manual intervention, making it adaptable to any new dataset. In the context of the MICCAI'17 MM-WHS (International Conference on Medical Image Computing and Computer Assisted Intervention, 2017, Multi-Modality Whole Heart Segmentation) challenge (Zhuang & Shen, 2016), several state-of-the-art heart segmentation models emerged and were summarized (Zhuang et al., 2019). For example, Yang, Bian, Yu, Ni, and Heng (2017) proposed a modified 3D U-Net with a hybrid loss to address the potential imbalance among different cardiac sub-structures, thereby improving the overall heart segmentation performance. Hybrid loss function is designed in a spatial-aware manner, which uses the ratio of the number of voxels belonging to the heart versus the total number of voxels as a weighting factor in the loss calculation. This accounts for the imbalance of the proportion of background voxels in the cropped image, which helps to reduce the interference of a large number of background voxels on model training. Meanwhile, two-stage methods are also gaining increasing attention in cardiac image analysis. Wang, MacGillivray, Macnaught, Yang, and Newby (2018) introduced a two-stage modified U-Net architecture that simultaneously detects a Region of Interest (RoI) from the full volume and segments the voxels at the original resolution. Payer, Štern, Bischof, and Urschler (2017) employed two CNNs (Convolutional Neural Networks) in an end-to-end manner, including a location CNN and a segmentation CNN, for whole heart segmentation. The location CNN first localizes the center of all heart substructures, aiding the subsequent segmentation CNN in focusing on the heart region. This spatial-aware model architecture also aims to address the challenges posed by the small proportion of heart voxels relative to the background. On the other hands, some methods incorporate the boundaries information into the network. For instance, a study (Park & Chung, 2022) introduces the shape-aware contour attentions to improve the overall segmentation. However, current methods for heart segmentation fail to consider the interrelationship among different heart sub-structures. This limitation results in poor EM segmentation results since the boundaries of EM are defined by the relative positions of certain important heart sub-structures (LV, Myo, etc.). The proposed methods similarly draw upon these spatially aware

design principles, including the two-stage approach and Hybrid loss function, in order to achieve high-precision heart segmentation despite the inherent difficulties of this task.

3. Dataset

The proposed dataset for HOCM segmentation comprises a total of 40 CT volumes obtained from HOCM patients who underwent septal myectomy surgery at Guangdong Provincial People's Hospital between January 6, 2020, and February 5, 2021. All the CT scans were captured using the Siemens Somatom Definition Flash CT scanner. The scanning was performed prospectively with a data acquisition window of 75 ms and a rack rotation time of 0.28 s/r. The scanning direction was from foot to head, covering the thorax entrance to 2.0 cm below the diaphragm. The acquired data have received necessary approvals from institutional ethics committees. The patients in the dataset had a mean age of 50.8 ± 16.9 and were diagnosed with HOCM, undergoing septal myectomy surgery. Prior to the CT scans, all patients were initially screened using B-ultrasound, and detailed surgical planning was carried out based on the diagnosis of HOCM. Consequently, the dataset does not include CT volumes from healthy patients. Each original CT volume in the dataset contains 277 to 842 slices, with an original resolution of 512 × 512 pixels per slice. EM labeling (including EM and RM labeling) is a crucial step in dataset preparation, and its definition is as follows: EM typically starts at the nadir of the right cusp, 5 mm below the aortic valve, and extends leftward towards the mitral anterior commissure, deepening and lengthening towards the apex of the heart. In the mid-ventricle, the resection extends rightward to the mid-inferoseptum. To ensure complete obstruction removal and operation safety, the scope of EM needs to be precisely defined. Based on domestic and foreign research (Maron, Rowin, Casey, & Maron, 2016; Morrow & Brockenbrough, 1961; Schoendube et al., 1995) and our previous work (Ma et al., 2021), we provide a detailed summary of the EM scope as follows:

- The upper boundary is 5 mm below the aortic valve annulus;
- The left boundary extends from the anterolateral commissure of the mitral valve to the left anterior free wall at the mid-septum;
- The right boundary extends from the basal layer near the membranous septum to the posterior septum at the mid-septum;
- The lower boundary is the projection plane of the papillary muscle root on the interventricular septum.
- When the septum thickness is greater than or equal to 24 mm, 50% is resected; When the thickness is less than 24 mm, the portion exceeding 12 mm is resected.
- The myocardium segment with a distance greater than 15 mm from the right ventricle is identified as the myocardium to be resected.

As shown in Fig. 2, the overall process consists of three steps: preprocessing, pre-operative surgical planning, and post-operative evaluation. Preprocessing is to acquire ultrasound and CT images which are manually annotated and fused to obtain a 3D segmentation model with cardiac sub-structures. During ultrasound image acquisition, the 135-degree long-axis section of the middle esophagus is used, and the frame rate needs to be controlled to a relatively low value to insure the quality of the images. During CT data acquisition, images in both diastole and systole are obtained, and the left and right heart chambers are required to be clearly visualized or enhanced. In addition, the heart rates during the acquisition of both the two examinations needs to be almost the same to ensure the fusion quality. The annotation process is performed by experienced surgeons. In the pre-operative surgical planning stage, surgeons first determine the range of excised myocardium based on patients' heart 3D segmentation model and patients' clinical condition. At the same time, the range of excised myocardium is measured in 3D and marked in the 3D segmentation model. Then, pre-operative surgical operation simulation is performed through 3D printing and VR to further enhance surgical understanding and optimize surgical planning. In the post-operative evaluation stage, surgeons comprehensively evaluates the effect of this operation by comparing the operative field before and after the surgery, the actual excised myocardium and the predicted excised myocardium, which provides assistance for the patients' postoperative treatment plan.

Fig. 1(a), (b), and (c) illustrate three views of the example CT volumes, the corresponding ground truth, and the 3D reconstruction results in the proposed dataset, respectively. It can be observed that there is no clear distinction between the EM and RM regions, as indicated by the red rectangle. The experienced surgeons also labeled five additional less-important surgery-related cardiac sub-structures, including the left ventricle (LV), right ventricle (RV), left atrium (LA), right atrium (RA), and aorta (Ao), for potential research on hypertrophic obstructive cardiomyopathy (HOCM). Note that other cardiac sub-structures, such as the superior vena cava, inferior vena cava, pulmonary veins, and pulmonary artery, were not labeled, as they require significant effort to annotate but provide limited benefit for surgical planning. Fig. 1(d) shows the 3D views of the EM boundaries of the segmentation results, providing an intuitive understanding of the spatial relationship between EM, RM, LV, RV, and the entire myocardium (Myo), which includes both EM and RM.

4. Method

4.1. Two-stage coarse-to-fine training framework

The overall structure of HOCM-Net is illustrated in Fig. 3. HOCM-Net consists of three components: a coarse stage (whole heart segmentation), a structural prior feature fusion module, and a fine stage. In the coarse stage, the aim is to locate the entire heart within the CT volume through whole heart segmentation. During the structural prior feature fusion, we initially extract the Region of Interest (RoI) image, which represents the heart's location, from the original CT volumes. Subsequently, we incorporate the RoI predictions (segmentation of LV, RV, and Myo) into the RoI image as additional input for the fine stage. Finally, the fine stage concentrates on the heart region and performs EM segmentation using the RoI, aided by supplementary input comprising the LV, RV, and Myo segmentation.

For clarity, The cardiac CT image consists of the voxels for background and the voxels for $\mathcal{K}$ kinds of cardiac sub-structures. We denote the proposed HOCM image dataset as $\mathcal{S}$, and $(\mathbf{X}, \mathbf{Y}) \in \mathcal{S}$. $\mathbf{X}$ is the original image, where $\mathbf{X} \in \mathbb{R}^{\mathcal{W}} \times \mathbb{R}^{\mathcal{H}} \times \mathbb{R}^{\mathcal{C}}$, and $\mathcal{W}, \mathcal{H}, \mathcal{C}$ represents the width, height, and number of channels in the image. Meanwhile, $\mathbf{Y}$ is the corresponding ground truth, where $\mathbf{Y} \in \{0,1\}^{\mathcal{K}+1}$. The dataset $\mathcal{S}$ is randomly divided into 2 disjoint folds, $(\mathbf{X_T}, \mathbf{Y_T}) \in \mathcal{S_T}$ and $(\mathbf{X_E}, \mathbf{Y_E}) \in \mathcal{S_E}$, to train and evaluate the proposed method and existing state-of-the-art methods. Furthermore, the model used in coarse stage and the fine stage are denoted as $\mathbf{\Lambda_C}$ and $\mathbf{\Lambda_F}$, respectively. The 3D version of pyramid input and 3D version of deep supervision are denoted as $\mathbf{\Phi_p}$ and $\mathbf{\Phi_d}$, respectively. The construction of the model in these two stages depends on 3D coarse-to-fine feature fusion module, and its detailed description will be shown in Section 4.3.

The coarse stage aims to determine the position of the entire heart in the CT volumes, which are then cropped for further image processing. This process reduces the voxel count of the background, thus addressing the potential imbalance issue between target voxels and the background. In addition to EM and RM, the other five cardiac sub-structures (LV, RV, LA, RA, and Ao) are also roughly segmented, aiding in more accurate localization of the entire heart. In the coarse stage, the training process of the coarse stage model $\mathbf{\Lambda_C}$ can be formulated as follow,

$$\boldsymbol{\Theta}_{\mathbf{C}}^{*} \leftarrow \underset{(\mathbf{X_T},\mathbf{Y_T})\in\mathcal{S_T},\boldsymbol{\Theta}_{\mathbf{C}}}{\arg\max} \log \boldsymbol{\Lambda}_{\mathbf{C}}(\mathbf{Y_T}|\mathbf{X_T}, \mathbf{\Phi_p}(\mathbf{X}); \boldsymbol{\Theta}_{\mathbf{C}}), \tag{1}$$

where $\mathbf{\Phi_p}(\mathbf{X})$ is the process of pyramid input, which is described in Section 4.2. Then, we can get the inference result $\mathbf{y_{T,C}}$ of the $\boldsymbol{\Lambda}_{\mathbf{C}}$, which

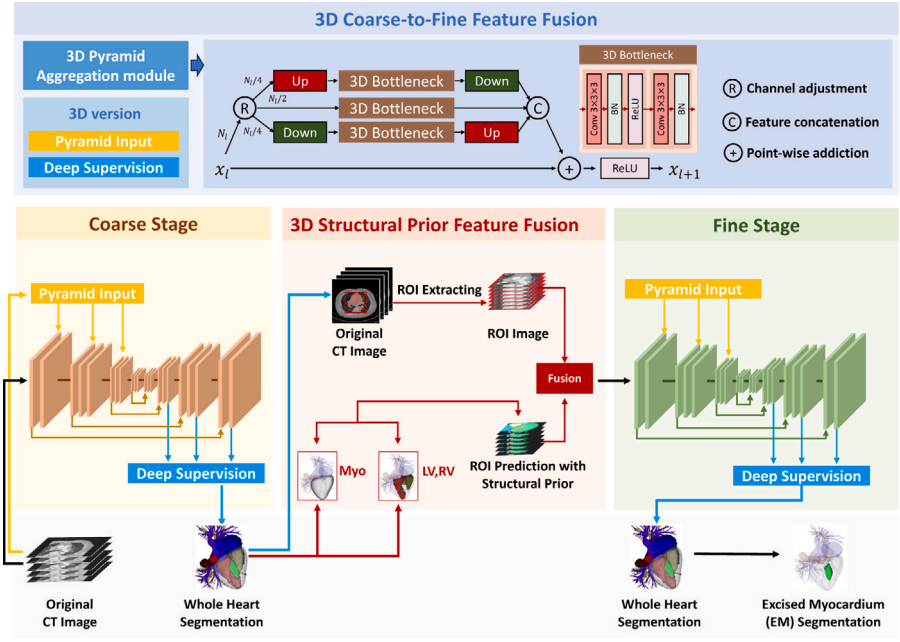


Fig. 3. Schematic illustration of the proposed HOCM-Net. HOCM-Net consists of two stages, the coarse stage and the fine stage. The basic structure of the network in both stages is 3D coarse-to-fine feature fusion module, which employs two auxiliary scale features (one feature at a higher scale, one feature at a lower scale) to enhance the current scale feature at each layer and thus to improve the overall segmentation. The coarse stage and the fine stage are connected by the 3D structural prior feature fusion module, which combines the structural information of left ventricle (LV), right ventricle (RV), and Myo segmentation result in the coarse stage to refine the EM segmentation in the fine stage.

can be formulated as, $\mathbf{y}_{\mathbf{T,C}} \leftarrow \boldsymbol{\Lambda}_{\mathbf{C}}(\mathbf{X}_{\mathbf{T}}, \boldsymbol{\Phi}_{\mathbf{p}}(\mathbf{X}_{\mathbf{T}}); \boldsymbol{\Theta}^{*}_{\mathbf{C}}), \mathbf{X}_{\mathbf{T}} \in \mathcal{S}_{\mathbf{T}}$. Next, we denote the RoI parameter determination function and RoI extraction function as $\boldsymbol{\delta}_{\mathbf{RoI}}$ and $\boldsymbol{\eta}_{\mathbf{RoI}}$. From this, we can get the first input of the fine stage for training $\mathbf{X}_{\mathbf{T,RoI}}$.

$$\mathbf{X}_{\mathbf{T,RoI}}, \mathbf{Y}_{\mathbf{T,RoI}} \leftarrow \boldsymbol{\eta}_{\mathbf{RoI}}(\mathbf{X}_{\mathbf{T}}, \mathbf{Y}_{\mathbf{T}}; \mathcal{W}^{*}, \mathcal{H}^{*}, \mathcal{C}^{*}), (\mathbf{X}_{\mathbf{T}}, \mathbf{Y}_{\mathbf{T}}) \in \mathcal{S}_{\mathbf{T}}, \tag{2}$$

where $\mathcal{W}^{*}, \mathcal{H}^{*}, \mathcal{C}^{*} \leftarrow \boldsymbol{\delta}_{\mathbf{RoI}}(\mathbf{y}_{\mathbf{T,C}})$. The second input of the fine stage is denoted as $\mathbf{y}^{*}_{\mathbf{T,C}}$, containing the predicted results of the coarse stage that are beneficial to the fine stage segmentation (Myo, LV, RV). Similarly, $\mathbf{Y}_{\mathbf{T}}, \mathbf{y}^{*}_{\mathbf{T,C}}$ are also cropped with the same parameter of $\mathbf{X}_{\mathbf{T,RoI}}$, and got $\mathbf{Y}_{\mathbf{T,RoI}}$ and $\mathbf{y}^{*}_{\mathbf{T,C,RoI}}$.

The purpose of the fine stage is to integrate the structural information obtained in the coarse stage results in order to enhance EM segmentation. The input feature maps consist of the structural prior feature fusion outcomes of the cropped original images and the cropped segmentation predictions of LV, RV, and Myo from the coarse stage. The segmentation of LV, RV, and Myo assists in guiding the fine stage segmentation to focus on the Myo region, thereby improving EM segmentation. In the fine stage, the training process of the coarse stage model $\boldsymbol{\Lambda}_{\mathbf{F}}$ can be formulated as follow,

$$\boldsymbol{\Theta}^{*}_{\mathbf{F}} \leftarrow \underset{(\mathbf{X}_{\mathbf{T,RoI}}, \mathbf{Y}_{\mathbf{T,RoI}}) \in \mathcal{S}_{\mathbf{T}}, \boldsymbol{\Theta}_{\mathbf{F}}}{\arg\max} \log \boldsymbol{\Lambda}_{\mathbf{F}}(\mathbf{Y}_{\mathbf{T,RoI}} | \mathbf{X}_{\boldsymbol{\Psi}}, \boldsymbol{\Phi}_{\mathbf{p}}(\mathbf{X}_{\mathbf{T,RoI}}); \boldsymbol{\Theta}_{\mathbf{F}}), \tag{3}$$

where $\mathbf{X}_{\boldsymbol{\Psi}}$ is the 3D structural Prior Feature fused result, which is formulated as, $\mathbf{X}_{\boldsymbol{\Psi}} \leftarrow \Psi(\mathbf{X}_{\mathbf{T,RoI}}, \mathbf{y}^{*}_{\mathbf{T,C,RoI}})$. $\Psi()$ will be described in Section 4.3.

Then, we can get the final inference result $\mathbf{y}_{\mathbf{E,F}}$ of the $\boldsymbol{\Lambda}_{\mathbf{F}}$, which can be formulated as, $\mathbf{y}_{\mathbf{E,F}} \leftarrow \boldsymbol{\Lambda}_{\mathbf{F}}(\mathbf{X}_{\mathbf{E,RoI}}, \boldsymbol{\Phi}_{\mathbf{p}}(\mathbf{X}_{\mathbf{E,RoI}}), \mathbf{y}^{*}_{\mathbf{C,RoI}}; \boldsymbol{\Theta}^{*}_{\mathbf{F}})$, where the definition of $\mathbf{X}_{\mathbf{E,RoI}}$ is similar to $\mathbf{X}_{\mathbf{T,RoI}}$, which is cropped by the same parameter with $\mathbf{X}_{\mathbf{T,RoI}}$.

4.2. 3D coarse-to-fine feature fusion module

In this paper, we propose a new 3D convolution structure (pyramid-scale aggregation) for coarse-to-fine feature fusion, and propose 3D versions of two 2D convolution methods (pyramid input Zhang et al., 2021, deep supervision Yang et al., 2017) to further promote the coarse-to-fine feature fusion of 3D convolution. Pyramid-scale aggregation (3D-PAM) is utilized as the fundamental module of each layer to integrate coarse-to-fine contextual information. Inspired by ResNet (He, Zhang, Ren, & Sun, 2016), 3D-PAM employs two associated branches (higher scale and lower scale) to assist the main branch (current scale), which follows the original residual learning path in ResNet. To minimize potential computational cost increments, the channels of the main branch are halved, while those of the associated branches are reduced to one-fourth. After the scale adjustment, each branch utilizes the 3D version of the bottleneck module from ResNet (He et al., 2016) to concurrently learn features, replacing 2D convolutions with 3D convolutions. In the 3D bottleneck module, two $3 \times 3 \times 3$ convolutions are employed to learn feature maps. The first convolution is followed by batch normalization (BN) and rectified linear unit (ReLU), while the second convolution is only followed by BN. The output feature maps of the associated branches are scaled to the same level as the main branch and concatenated with the output feature maps of the main branch.

Furthermore, to enhance segmentation performance, we propose various implementation optimizations, including pyramid input and deep supervision, in both the coarse and fine stages. Pyramid input resizes the original CT volumes to corresponding scales, enabling the feature maps in the intermediate layers of the network to incorporate coarse information and enhance segmentation. To minimize the need for network revisions, each pyramid fused result is followed by a $3 \times 3 \times 3$ convolutional layer and ReLU, adjusted to match the original number of feature map channels. Notably, the pyramid input in the fine stage also serves as the structural prior feature fusion result. Deep supervision optimizes the intermediate feature maps using the corresponding scale's label, speeding up the training process and improving overall segmentation.

4.3. 3D structural prior feature fusion module

In this section, we introduce the 3D structural prior feature fusion module, which integrates prior medical knowledge to enhance the EM segmentation. Particularly, based on the definition of EM and our experience of manual segmentation (Ma et al., 2021), we find that there

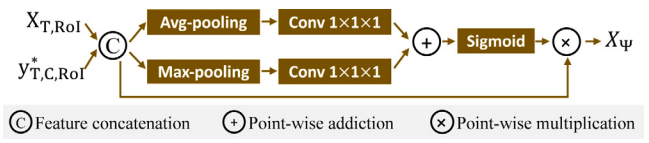


Fig. 4. Schematic illustration of the proposed structural prior feature fusion. 3D structural prior feature fusion module aims at aggregating RoI predictions (the cropped segmentation results of LV, RV, and Myo) to the RoI images (the cropped CT volumes in the coarse stage) as extra input for the EM segmentation in the fine stage.

exist explicit position relationships between EM, RM, LV, and RV. First, the target area including EM is roughly located between LV and RV, and at the center of the combination of LV and RV. Second, around the target area, LV, EM, RM, and RV are almost perfectly stacked in a sandwich manner. Thus, the location of LV, RV, and RM are important for the segmentation of EM, which is also confirmed in the boundary definition of EM. The aggregation of the context information of LV, RV, and RM to enhance EM segmentation is illustrated in Fig. 4. To integrate the above prior medical knowledge, the segmentation of LV, RV, and Myo at the coarse stage is extracted as the extra input. In this way, the fine stage can get an idea of the rough location of LV, RV, and Myo, which could then help locate EM. RoI images x_{vol} and RoI predictions x_{pred} denote the cropped results of the original CT volumes and predictions in the coarse stage, respectively. The feature concatenation of RoI image and RoI predictions is formulated as follows, where H denotes feature concatenation and is followed by a $3 \times 3 \times 3$ convolutional layer with batch normalize and ReLU. Then the fused result is then processed by a channel attention module (Hu, Shen, & Sun, 2018) as follows,

$$\mathbf{x_{fusion}} = \sigma(\mathbf{M}(\mathbf{\Phi_{Avg}}(\mathbf{x_{concat}}))) \otimes \mathbf{x_{concat}}, \tag{4}$$

which is used for the enhancement of the integration by exploiting the inter-channel relationship of features. σ denotes the Sigmoid function, while $\mathbf{M}$ is the fully connected layer, and $\otimes$ denotes element-wise multiplication. Note that average-pooling $\mathbf{\Phi_{Avg}}$ is used to squeeze the spatial dimensions of the input feature maps to aggregate channel information.

5. Experiments

5.1. Experimental setup

In this paper, we implemented the proposed methods on the TensorFlow platform with 1.3 version using two NVIDIA GeForce GTX 4090 GPUs, each with 24 GB of memory. The dataset contains 40 instances from the HOCM patients, which are randomly divided into 2 disjoint folds to train and test the proposed method and existing state-of-the-art methods. To account for memory restrictions imposed by volumetric inputs and 3D convolution layers, we utilized a mini-batch size of 1. The input size of the network was set to $96 \times 96 \times 96$. Besides, we chose the Adam optimizer with an initial learning rate of 0.001. Meanwhile, Considering the imbalance between EM and RM, we chose hybrid loss (Yang et al., 2017) to optimize the network. Hybrid loss takes the ratio of the pixels belonging to the target structure to all the pixels in the input image as the weight of the loss function, which helps to suppress the imbalance caused by the input images with a small proportion of the target structure. Particularly, we introduced image normalization, random cropping, and random rotation with an angle range of $[-20°,20°]$ to augment data to alleviate the potential overfitting. Moreover, to ensure fair comparisons, we used the ROI images generated from the coarse stage to train state-of-the-art networks in our experiments.

To showcase the advantages of the proposed HOCM-Net in the automatic segmentation of EM, we conducted a comparison with six CNN approaches:

- **3D U-Net** (Çiçek et al., 2016): Based on the classic U-Net, 3D U-Net extends the previous architecture by replacing all 2D operations with their 3D counterparts. It is the most widely used method in 3D medical image segmentation.
- **T3D U-Net** (Wang et al., 2018): T3D U-Net is a two-stage approach that modifies the original U-Net. It simultaneously learns to detect an interesting region from the whole image and classify voxels while maintaining the original resolution. This approach demonstrates significant improvements in whole heart segmentation.
- **ML-CNN** (Payer et al., 2017): ML-CNN proposes two FCNs for automatic whole heart segmentation from CT and MRI volumes. The first CNN localizes the heart structures, enabling the second CNN to focus on this region. ML-CNN outperforms other participants in CT test sets of the MICCAI'17 MM-WHS (Zhuang & Shen, 2016) challenge.
- **Hybrid-Net** (Yang et al., 2017): Hybrid-Net combines the FCN with 3D operators, transfer learning, and a deep supervision mechanism to distill 3D contextual information and improve whole heart segmentation. Additionally, it introduces a new hybrid loss to address the serious imbalance among cardiac substructures. Hybrid-Net achieved first place in the MICCAI'17 MM-WHS (Zhuang & Shen, 2016) challenge.
- **MDFN** (Ye, Wang, Zhang, & Wang, 2019): Multi-depth fusion network (MDFN) applies multi-depth fusion to the original 3D U-Net, enabling the fusion of multi-scale context information for better extraction of context information. It has shown promising results in whole heart segmentation.
- **nnU-Net** (Isensee et al., 2021): No-new-U-Net (nnU-Net) proposes a medical image segmentation framework that automatically adjusts all hyperparameters according to the attributes of a given dataset without requiring manual intervention. nnU-Net is adaptable to most new datasets and is widely utilized in various medical image analyses.

The basic architecture used in the coarse stage is a modified version of the 3D U-Net. It consists of sets of 32, 64, 128, 256, and 512 convolutional filters in each encoder and decoder path. To mitigate the influence of initial mistakes from the network in the coarse stage on the network in the fine stage, we utilize a step-by-step training strategy inspired by the training process of Faster R-CNN (Ren, He, Girshick, & Sun, 2015). Our training process consists of three steps. First, we independently train the network in the coarse stage to obtain coarse segmentation of the cardiac sub-structures. Next, we freeze the network in the coarse stage and train the network in the fine stage to achieve fine segmentation of the cardiac sub-structures. Finally, we fine-tune the entire two-stage architecture and obtain the final EM segmentation result from the fine segmentation.

5.2. Evaluation metric

In this paper, we introduce three metrics for evaluation: the Dice similarity coefficient (Dice), the average symmetric surface distance (ASD), and the 95 percent Hausdorff distance (95HD). The dice metric accesses the spatial overlap between segmentation prediction and the ground truth, which can be formulated as

$$D(G,S) = 2|G \cap S|/(|G| + |S|), \tag{5}$$

where G and S represent ground truth and prediction segmentation, respectively. ASD averages all distances from the voxels on the surface of the segmentation prediction to the ground truth, which is defined as,

$$ASD(G,S) = \frac{1}{2}\left(\frac{1}{\|G\|}\sum_{x\in G}\inf_{y\in S}\|x-y\| + \frac{1}{\|S\|}\sum_{y\in S}\inf_{x\in G}\|x-y\|\right). \tag{6}$$

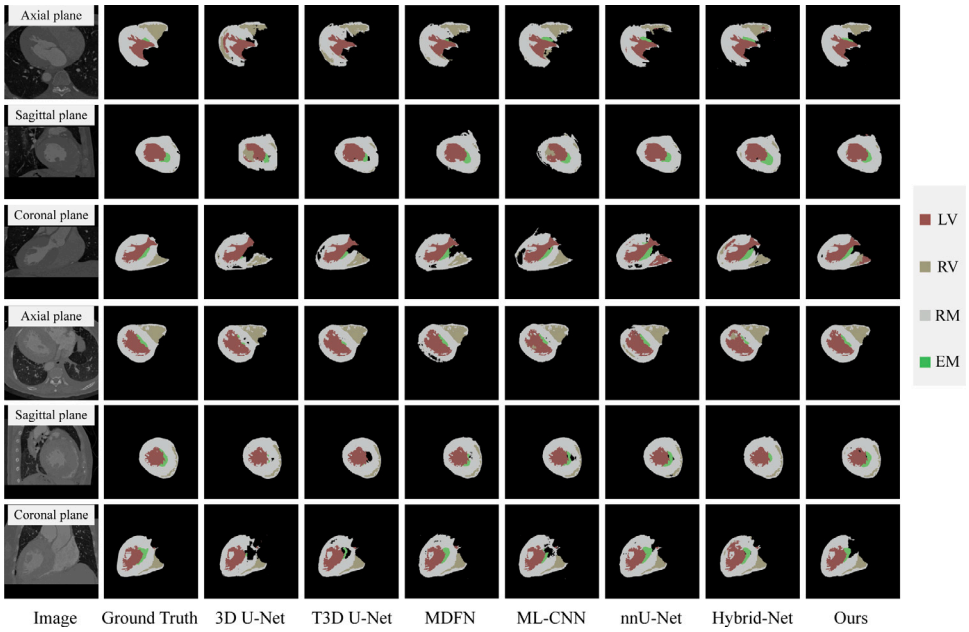


Fig. 5. 2D visualization of the segmentation results of HOCM-Net and existing methods (3D U-Net, T3D U-Net, MDFN, ML-CNN, nnU-Net, and Hybrid-Net). Three planes (axial plane, sagittal plane, coronal plane) are adopted for visualization, and the four important classes including LV, RV, RM, and EM are shown.

Table 2
EM segmentation performance comparison and statistic analysis of HOCM-Net and existing methods (3D U-Net, T3D U-Net, MDFN, ML-CNN, nnU-Net, and Hybrid-Net).

Method	Dice	95HD (mm)	ASD (mm)	p-value
3D U-Net (Çiçek et al., 2016)	0.46 ± 0.21	22.22 ± 11.54	2.80 ± 1.01	<0.001
T3D U-Net (Wang et al., 2018)	0.59 ± 0.14	13.41 ± 5.32	2.89 ± 0.97	<0.001
MDFN (Ye et al., 2019)	0.67 ± 0.10	12.17 ± 3.16	2.91 ± 0.75	<0.001
nnU-Net (Isensee et al., 2021)	0.68 ± 0.08	11.45 ± 4.12	2.71 ± 0.82	<0.001
ML-CNN (Payer et al., 2017)	0.68 ± 0.12	11.39 ± 3.67	2.82 ± 0.92	<0.001
Hybrid-Net (Yang et al., 2017)	0.70 ± 0.09	10.13 ± 3.21	2.74 ± 1.12	<0.005
Ours	**0.75 ± 0.06**	**8.13 ± 2.87**	**2.26 ± 0.97**	–

Hausdorff distance evaluates the shape similarity of the segmentation results, which is formulated as,

$$H(G,S) = \max\left\{\sup_{x\in G}\inf_{y\in S}\|x-y\|, \sup_{y\in S}\inf_{x\in G}\|x-y\|\right\}. \quad (7)$$

Meanwhile, 95HD is defined as the 0.95-quantile partial distance between the two sets, which is an outlier-robust measure based on the Hausdorff distance.

5.3. Quantitative results

We compared our proposed method HOCM-Net with six existing state-of-the-art models including 3D U-Net (Çiçek et al., 2016), T3D U-Net (Wang et al., 2018), MDFN (Ye et al., 2019), ML-CNN (Payer et al., 2017), and Hybrid-Net (Yang et al., 2017), and Table 2 summarizes the experimental results of EM including Dice score, 95HD, and ASD. 3D U-Net failed to obtain good results with a Dice score of 0.46, a 95HD performance of 22.22, and an ASD performance of 2.80. Some methods like T3D U-Net and ML-CNN work in a two-stage way (extracts RoIs first, and then performs segmentation with largely reduced redundant information), which usually have much better performance. For example, T3D U-Net obtains 0.59 on Dice score, 13.41 on 95HD, and 2.89 on ASD, while ML-CNN (Payer et al., 2017) achieves 0.68 on Dice score, 11.39 on 95HD, and 2.82 on ASD. Meanwhile, by adjusting the hyper-parameters automatically, nnU-Net (Isensee et al., 2021) obtains 0.68 on Dice score, 11.45 on 95HD, and 2.71 on ASD, which is comparable with two-stage methods. Some methods like MDFN and Hybrid-Net introduced a volume size weighted loss function combining the cross entropy loss and Dice loss to alleviate the imbalance problem among all cardiac sub-structures in the images, which can further improve the segmentation results. For example, MDFN (Ye et al., 2019) obtains 0.67 on Dice score, 12.17 on 95HD, and 2.91 on ASD, while Hybrid-Net obtains 0.70 on Dice score, 10.13 on 95HD, and 2.74 on ASD. Benefiting from pyramid input, deep supervision, 3D structural prior feature fusion, and 3D-PAM, our proposed method HOCM-Net surpasses previous efforts with a state-of-the-art performance of 0.75 (improved by 5.0%) on Dice score, 8.13 (improved by 17.9%) on 95HD, and 2.26 (improved by 17.5%) on ASD. Statistical analyses show that our HOCM-Net achieves significantly higher performance on all three metrics than existing works.

5.4. Qualitative results

Figs. 5 and 6 show the 2D and 3D visual comparison between HOCM-Net and six state-of-the-art methods, including 3D U-Net, T3D U-Net, MDFN, ML-CNN, nnU-Net, and Hybrid-Net on our proposed datasets.

Fig. 5 displays 2D segmentation results including three planes (from top to bottom: axial plane, sagittal plane, coronal plane) with four classes including LV (dark red), RV (dark yellow), Myo (gray), and EM (green) from two patients with HOCM disease. We can notice that our proposed HOCM-Net achieves consistently higher segmentation accuracy than existing methods in all rows except Row 6, which suffers a big missing on Myo. This may be due to the large variation of heart structures, especially in HOCM patients. In terms of continuity, we can see that hybrid-Net and our proposed HOCM-Net enjoy a good continuity, while 3D U-Net and T3D U-Net suffer some missing and

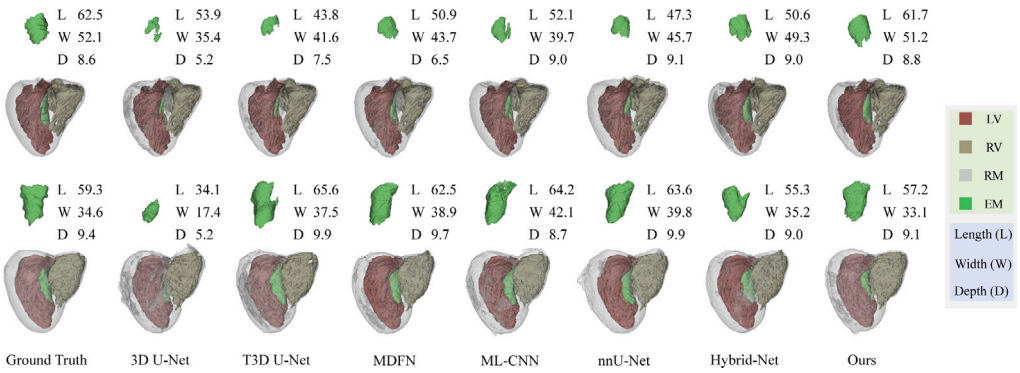


Fig. 6. 3D visualization of EM segmentation results of HOCM-Net and existing methods (3D U-Net, T3D U-Net, MDFN, ML-CNN, nnU-Net, and Hybrid-Net). All the numbers are with millimeter (mm) units.

Table 3
Surgeons' EM segmentation evaluation of HOCM-Net and Hybrid-Net the existing method with the highest accuracy in Table 2. Experienced surgeons made the comparison and determined which method has a better accuracy considering the clinical requirement.

Surgeon	6 difficult cases			6 simple cases		
	Ours	Hybrid-Net	Equal	Ours	Hybrid-Net	Equal
Surgeon 1	5	0	1	4	0	2
Surgeon 2	4	1	1	3	1	2
Surgeon 3	6	0	0	5	1	0
Majority	6	0	0	4	1	1

even appear to have some segmentation contradictions, especially on the segmentation of Myo (Row 3, Row 6). A potential reason is the sliding window configuration that makes some parts of RM and EM locate in an abnormal position in some inputs. In this way, some voxels of EM and RM could not be well recognized considering the large variety of heart. A piece of obvious evidence is the segmentation results of 3D U-Net shown in Row 6 where exists obvious discontinuous segmentation for Myo.

Fig. 6 illustrates the 3D reconstructed segmentation results and the 3D size of those segmentation results. In terms of the 3D size of the EM, it is evident that although Hybrid-Net occasionally outperforms our approach (as seen in the width of the second case), our HOCM-Net (2.7%) surpasses all existing methods, including 3D U-Net (37.0%), T3D U-Net (14.0%), MDFN (13.4%), ML-CNN (13.8%), nnU-Net (11.7%), and Hybrid-Net (7.0%). We attribute these improvements to the incorporation of 3D structural prior feature fusion and the aggregation of Myo, LV, and RV segmentations, which play a crucial role in defining the boundaries. In Figs. 5 and 6, it is noticeable that the segmentation accuracy of LV and RV generally appears to be higher than that of EM and RM. This discrepancy may be attributed to the varying shapes and class imbalance issues. RM exhibits a more diverse shape with two cavities compared to LV and RV, which are compact and lack internal cavities. Furthermore, when compared to LV and RV, which encompass larger regions, EM consists of a much smaller region that suffers from severe imbalance problems.

5.5. Surgeons' evaluation results

Table 3 illustrates the evaluation results of experienced surgeons regarding the effectiveness of EM for surgical planning in HOCM. It is important to note that due to the workload of the surgeons, we only compared our HOCM-Net with Hybrid-Net (the model with the highest accuracy in Table 2). We carefully selected 6 simple cases and 6 difficult cases based on the quality of the original reconstructed images. Additionally, we provided the 3D size of EM from different models. To ensure a closer resemblance to real-life operations, the size measurements were based on the long and short axes of the heart, as determined by the surgeons. By examining the 3D segmentation results, the surgeons were able to assess the relative size and positional relationships, as well as specific details such as the width and thickness of the resection area in the 2D results. The surgeons evaluated the segmentation accuracy by considering both the 2D and 3D segmentation results, particularly emphasizing the 3D size measurements. Their assessment took into account the clinical requirements described in Section 3. The last row of the table represents the majority agreement among the three surgeons. Overall, the evaluation clearly demonstrates that our methods outperform Hybrid-Net by a significant margin. Notably, for the difficult cases, all surgeons agreed that HOCM-Net achieved superior segmentation in almost all 6 cases compared to Hybrid-Net. In only two cases, as evaluated by Surgeon 1 and Surgeon 2, the accuracy of the two methods was considered equal. Furthermore, there was only one case, as assessed by Surgeon 2, where Hybrid-Net exhibited better accuracy than HOCM-Net. Similar findings were observed for the 6 simple cases, albeit with a smaller performance margin. This could be attributed to the fact that both methods performed well in simple cases.

5.6. Ablation analysis

In this section, we perform an ablation analysis to showcase the effectiveness of 3D-PAM, 3D structural prior feature fusion, and pyramid input enhancement in HOCM-Net. The experimental results of EM are summarized in Table 4. Firstly, we evaluated the effectiveness of 3D-PAM on the backbone. By leveraging the aggregation of coarse-to-fine 3D context information from pyramid scales in each layer, the backbone model with 3D-PAM achieves improvements of 1% on Dice and 0.52 mm on 95HD, with a comparable result in ASD, which is only 0.03 worse. Secondly, we assessed the impact of the 3D structural prior feature fusion module. Ablation results demonstrate that 3D structural prior feature fusion significantly enhances the segmentation performance by 3% on Dice, 0.99 mm on 95HD, and 0.30 mm on ASD. Thirdly, the evaluation of pyramid input enhancement and deep pyramid supervision also reveals that HOCM-Net obtains certain improvements through the aforementioned two optimizations. Overall, the primary improvement stems from the 3D structural prior feature fusion, which aggregates the segmentation results of LV, RV, and Myo. 3D-PAM and the other two optimizations can also contribute to some improvements.

On the other hand, we also conduct an ablation analysis of the techniques used in the pre-processing and post-processing stages of the HOCM-Net, which is shown in Table 5. The results show that when the random cropping, random rotation, or removal of small components techniques are individually removed, the performance of the model decreases across all three metrics. This suggests that these pre-processing and post-processing techniques play an important role in the overall performance of the HOCM-Net and contribute to its effectiveness.

Table 4
Ablation analysis of 3D-PAM, 3D structural prior feature fusion, pyramid input, and deep supervision in HOCM-Net.

3D-PAM	3D Structural prior feature fusion	Pyramid input	Deep supervision	Dice	95HD	ASD
				0.70	10.68	2.65
✓				0.71	10.16	2.68
✓	✓			0.74	9.17	2.38
✓	✓	✓		0.74	9.06	2.31
✓	✓	✓	✓	0.75	8.13	2.26

Table 5
The ablation analysis of techniques used in pre-process and post-process.

Methods	Dice	95HD (mm)	ASD (mm)
HOCM-Net	0.75 ± 0.06	8.13 ± 2.87	2.26 ± 0.97
- w/o random crop	0.73 ± 0.09	8.46 ± 2.95	2.42 ± 0.92
- w/o random rotation	0.73 ± 0.10	8.57 ± 2.98	2.38 ± 0.99
- w/o remove small components	0.72 ± 0.13	8.89 ± 3.26	2.59 ± 1.13

5.7. Extended evaluation

In clinical practice, segmenting various cardiac sub-structures such as LV, RV, and Myo can provide surgeons with a better understanding of their relationship with EM. However, there is a moderate requirement for segmentation accuracy in this regard. In this section, we conduct an extended evaluation of these cardiac sub-structures. Table 6 presents the segmentation results for these sub-structures and evaluates them using the Dice metric. Regarding the different cardiac sub-structures, it is noticeable that LV and RV achieve significantly higher accuracy compared to LA, RA, and Ao. Several potential reasons can explain these observations. Firstly, in the second stage, the input consists of cropped results containing the entire Myo. As a result, LV and RV, which are wrapped in Myo, are located near the center of the input image, attracting more attention. Secondly, LV and RV have a fixed positional relationship with Myo, making them easier to learn. Thirdly, in cardiac CTs of HOCM patients, the scanning configuration sometimes excludes a portion of Ao and Ao arch, causing Ao to be divided into ascending Ao and descending Ao. This introduces a greater variation in the shape of Ao, making its segmentation relatively challenging. Fourthly, LA and RA are connected to other structures (e.g., RA is connected to both superior vena cava and inferior vena cava, while LA is connected to pulmonary veins), making the segmentation of their boundaries more difficult. Compared to existing methods, our approach not only achieves the highest accuracy in EM segmentation but also demonstrates much higher accuracy in all other sub-structures. One possible reason for this is that other sub-structures also benefit from structure fusion. By leveraging prior medical knowledge, particularly the approximate structure and position of LV, RV, and Myo, all sub-structures can be segmented more accurately.

5.8. Discussion

In this paper, we present a EM segmentation approach for the surgical planning of HOCM surgery. The accurate segmentation of EM remains challenging due to the lack of local discrimination between EM and RM, as well as class imbalance among cardiac sub-structures. To address this, we propose a two-stage network in a spatial-aware manner, which focuses the second network on the heart region predicted by the first stage, aiming to alleviate potential imbalance. Additionally, we utilize the location of LV, RV, and Myo to guide the segmentation of EM, as EM should be contained within Myo and close to LV. The network is equipped with 3D-PAM, which leverages 3D multi-scale context information to enhance segmentation performance. Furthermore, we introduce a pyramid input that provides coarse-level context information at multiple scales, and deep supervision that accelerates training and improves multi-scale target performance. One of the main challenges in automatic segmentation of cardiac sub-structures is the presence of massive redundant information and severe imbalance among them. For instance, scar quantization of Myo (Fahmy et al., 2020; Ingles et al., 2019) in patients with HOCM is challenging due to the small proportion of scars, resulting in serious imbalance. While the segmentation of ventricles shows promising results (Dice score of 0.82), scar segmentation performance is poor (Dice score of 0.57). Recent works have employed methods that extract target regions from original CT volumes. For example, ML-CNN uses a CNN to locate the heart region and directly segments the whole heart within that region to improve overall segmentation. Another approach, presented in Xu, Wu, and Feng (2018), embeds 3D U-Net into a modified Faster R-CNN network to locate and crop the heart in cardiac CT volumes before performing segmentation. The proposed methods leverage spatially-aware design principles, such as the two-stage approach and hybrid loss function, to overcome the inherent challenges and achieve high-precision heart segmentation. However, these methods separate location and segmentation, failing to fully utilize the relationships between different cardiac sub-structures. Furthermore, the segmentation of EM differs from other cardiac sub-structures as it is defined by its relative position with other sub-structures rather than having clear discrimination from RM. Aggregating feature maps from different branches for different tasks has shown to improve performance and training speed. For example, Mask scoring R-CNN (Huang, Huang, Gong, Huang, & Wang, 2019) fuses classification results with the segmentation network to enhance segmentation performance. In clinical practice, EM is considered a part of the original Myo, situated close to LV, and defined by the relative positions of other sub-structures. To leverage this knowledge, we propose a 3D structural prior feature fusion module that combines the location results from the first stage with the original cardiac CT volume in the second stage. This fusion module preserves the original information for detailed segmentation while aggregating previous results to enhance overall segmentation. Our experiments demonstrate that our proposed model outperforms current state-of-the-art heart segmentation models, which we attribute to its superior awareness of LV, RV, and Myo segmentation in the second stage, predicted in the first stage. In conclusion, our goal in segmenting the main cardiac sub-structures is twofold. Firstly, we aim to locate the heart by using more distinguishable sub-structures (LV and RV) to mitigate bias caused by class imbalance. Secondly, we aim to improve EM segmentation by incorporating structure information from related cardiac sub-structures (LV, RV, and Myo) as prior medical knowledge.

6. Conclusion

In this paper, we propose a 3D structural prior feature fusion-based network called HOCM-Net for EM Segmentation in surgical planning for HOCM. HOCM-Net incorporates prior medical knowledge to achieve accurate EM segmentation. Specifically, the network fuses the structural information from the segmentation of left ventricle (LV), right ventricle (RV), and myocardium (Myo) with the input image to aid in locating the boundary between EM and surrounding tissues. Additionally, HOCM-Net introduces a 3D version of the multi-scale feature aggregation module to combine coarse-to-fine contextual information and improve the segmentation process. To evaluate the performance, we have assembled the first HOCM segmentation dataset, which consists of 40 CT volumes. The dataset includes annotations for EM, right myocardium (RM), and five other cardiac sub-structures

Table 6
Extended evaluation of HOCM-Net with existing methods on seven cardiac sub-structures (LV, RV, LA, RA, Ao, EM, RM).

Method	LV	RV	LA	RA	Ao	EM	RM	Mean
3D U-Net (Çiçek et al., 2016)	0.71	0.65	0.63	0.52	0.54	0.46	0.64	0.61 ± 0.15
T3D U-Net (Wang et al., 2018)	0.80	0.74	0.71	0.55	0.63	0.59	0.68	0.68 ± 0.16
MDFN (Ye et al., 2019)	0.85	0.86	0.73	0.55	0.67	0.67	0.77	0.74 ± 0.12
ML-CNN (Payer et al., 2017)	0.86	0.85	0.78	0.60	0.72	0.68	0.75	0.75 ± 0.09
nnU-Net (Isensee et al., 2021)	0.85	0.86	0.75	0.63	0.71	0.68	0.77	0.75 ± 0.11
Hybrid-Net (Yang et al., 2017)	0.87	0.85	0.76	**0.67**	0.70	0.70	0.79	0.77 ± 0.08
HOCM-Net	**0.89**	**0.88**	**0.79**	0.64	**0.76**	**0.75**	**0.84**	**0.80 ± 0.07**

provided by experienced surgeons. We have conducted a variety of experiments, and the results demonstrate that HOCM-Net outperforms several state-of-the-art methods, highlighting its potential for clinical applications. Moreover, we have made the dataset publicly available to the research community, aiming to facilitate further investigations into this interesting and challenging problem.

CRediT authorship contribution statement

Jiawei Zhang: Conceptualization, Methodology, Software, Writing – original draft. **Xiaodong Wang:** Data curation. **Hailong Qiu:** Data curation. **Yanchun Zhang:** Project administration, Supervision. **Weihong Han:** Project administration, Supervision. **Jialin Wang:** Data curation. **Tianchen Wang:** Data curation. **Yiyu Shi:** Resource, Writing – review & editing. **Meiping Huang:** Resource, Writing – review & editing. **Jian Zhuang:** Resource, Writing – review & editing. **Huiming Guo:** Resource, Writing – review & editing. **Xiaowei Xu:** Project administration, Supervision.

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.

Data availability

Data will be made available on request.

Acknowledgments

This work and the collection of data of retrospective data on implied consent received Research Ethics Committee (REC) approval from Guangdong Provincial People's Hospital, Guangdong Academy of Medical Science under Protocol No. 2022-048-01. It complies with all relevant ethical regulations. Deidentification was performed in which all CT files are transformed into NIfTI format, and sensitive information of the patients including name, birther day, admission year, admission number, and CT number is removed. Only de-identified retrospective data were used for research, without the active involvement of patients.

This work was supported by the Natural Science Foundation of China (No. 62276071, No. 8210101340), Guangdong Special Support Program-Science and Technology Innovation Talent Project (No. 0620220211), the Science and Technology Planning Project of Guangdong Province, China (No. 2019B020230003), Guangdong Peak Project (No. DFJH201802), Guangzhou Science and Technology Planning Project (No. 202206010049), Guangdong Basic and Applied Basic Research Foundation (No. 2022A1515010157, No. 2022A1515011650), Guangzhou Science and Technology Planning Project (No. 202102080188).

References

Augusto, J. a. B., Davies, R. H., Bhuva, A. N., Knott, K. D., Seraphim, A., Alfarih, M., et al. (2021). Diagnosis and risk stratification in hypertrophic cardiomyopathy using machine learning wall thickness measurement: a comparison with human test-retest performance. *The Lancet Digital Health*, *3*(1), e20–e28.

Barker, J., Li, X., Khavandi, S., Koeckerling, D., Mavilakandy, A., Pepper, C., et al. (2022). Machine learning in sudden cardiac death risk prediction: a systematic review. *Europace*, *24*(11), 1777–1787.

Bartel, T., Rivard, A., Jimenez, A., Mestres, C. A., & Müller, S. (2018). Medical three-dimensional printing opens up new opportunities in cardiology and cardiac surgery. *European Heart Journal*, *39*(15), 1246–1254.

Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.-A., et al. (2018). Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved? *IEEE Transactions on Medical Imaging*, *37*(11), 2514–2525.

Çiçek, O., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention* (pp. 424–432). Springer.

Cui, H., Schaff, H. V., Lentz Carvalho, J., Nishimura, R. A., Geske, J. B., Dearani, J. A., et al. (2021). Myocardial histopathology in patients with obstructive hypertrophic cardiomyopathy. *Journal of the American College of Cardiology*, *77*(17), 2159–2170.

Fahmy, A. S., Neisius, U., Chan, R. H., Rowin, E. J., Manning, W. J., Maron, M. S., et al. (2020). Three-dimensional deep convolutional neural networks for automated myocardial scar quantification in hypertrophic cardiomyopathy: a multicenter multivendor study. *Radiology*, *294*(1), 52–60.

Farooqi, K. M., & Mahmood, F. (2018). Innovations in preoperative planning: insights into another dimension using 3D printing for cardiac disease. *Journal of Cardiothoracic and Vascular Anesthesia*, *32*(4), 1937–1945.

Geske, J. B., Sorajja, P., Ommen, S. R., & Nishimura, R. A. (2009). Left ventricular outflow tract gradient variability in hypertrophic cardiomyopathy. *Clinical Cardiology: An International Indexed and Peer-Reviewed Journal for Advances in the Treatment of Cardiovascular Disease*, *32*(7), 397–402.

Hamatani, Y., Amaki, M., Kanzaki, H., Yamashita, K., Nakashima, Y., Shibata, A., et al. (2017). Contrast-enhanced computed tomography with myocardial three-dimensional printing can guide treatment in symptomatic hypertrophic obstructive cardiomyopathy. *ESC Heart Failure*, *4*(4), 665–669.

Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H. R., & Xu, D. (2021). Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In *International MICCAI brainlesion workshop* (pp. 272–284). Springer.

Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., et al. (2022). Unetr: Transformers for 3d medical image segmentation. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 574–584).

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

Henn, M. C., & Mokadam, N. A. (2021). Three-dimensional printing to plan intracardiac operations. *JTCVS Techniques*, *9*, 101.

Hodges, K., Rivas, C. G., Aguilera, J., Borden, R., Alashi, A., Blackstone, E. H., et al. (2019). Surgical management of left ventricular outflow tract obstruction in a specialized hypertrophic obstructive cardiomyopathy center. *The Journal of Thoracic and Cardiovascular Surgery*, *157*(6), 2289–2299.

Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7132–7141).

Huang, Z., Huang, L., Gong, Y., Huang, C., & Wang, X. (2019). Mask scoring r-cnn. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 6409–6418).

Iacovoni, A., Spirito, P., Simon, C., Iascone, M., Di Dedda, G., De Filippo, P., et al. (2012). A contemporary European experience with surgical septal myectomy in hypertrophic cardiomyopathy. *European Heart Journal*, *33*(16), 2080–2087.

Ingles, J., Goldstein, J., Thaxton, C., Caleshu, C., Corty, E. W., Crowley, S. B., et al. (2019). Evaluating the clinical validity of hypertrophic cardiomyopathy genes. *Circulation: Genomic and Precision Medicine*, *12*(2), Article e002460.

Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., & Maier-Hein, K. H. (2021). NnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, *18*(2), 203–211.

Jung, C., Wolff, G., Wernly, B., Bruno, R. R., Franz, M., Schulze, P. C., et al. (2022). Virtual and augmented reality in cardiovascular care: state-of-the-art and future perspectives. *Cardiovascular Imaging*, *15*(3), 519–532.

Kamp, N. J., Chery, G., Kosinski, A. S., Desai, M. Y., Wazni, O., Schmidler, G. S., et al. (2021). Risk stratification using late gadolinium enhancement on cardiac magnetic resonance imaging in patients with hypertrophic cardiomyopathy: a systematic review and meta-analysis. *Progress in Cardiovascular Diseases*, *66*, 10–16.

Khurshid, S., Lazarte, J., Pirruccello, J. P., Weng, L.-C., Choi, S. H., Hall, A. W., et al. (2023). Clinical and genetic associations of deep learning-derived cardiac magnetic resonance-based left ventricular mass. *Nature Communications*, *14*(1), 1558.

Kim, L. K., Swaminathan, R. V., Looser, P., Minutello, R. M., Wong, S. C., Bergman, G., et al. (2016). Hospital volume outcomes after septal myectomy and alcohol septal ablation for treatment of obstructive hypertrophic cardiomyopathy: US Nationwide Inpatient Database, 2003–2011. *JAMA Cardiology*, *1*(3), 324–332.

Ma, J., Liu, J., Yuan, H., Tang, Y., Qiu, H., Zhuang, J., et al. (2021). Two-port thoracoscopic myectomy for hypertrophic cardiomyopathy with three-dimensional printing. *The Annals of Thoracic Surgery*, *111*(3), e165–e168.

Mancio, J., Pashakhanloo, F., El-Rewaidy, H., Jang, J., Joshi, G., Csecs, I., et al. (2022). Machine learning phenotyping of scarred myocardium from cine in hypertrophic cardiomyopathy. *European Heart Journal-Cardiovascular Imaging*, *23*(4), 532–542.

Maron, B. J. (2018). Clinical course and management of hypertrophic cardiomyopathy. *New England Journal of Medicine*, *379*(7), 655–668.

Maron, B. J., Desai, M. Y., Nishimura, R. A., Spirito, P., Rakowski, H., Towbin, J. A., et al. (2022). Diagnosis and evaluation of hypertrophic cardiomyopathy: JACC state-of-the-art review. *Journal of the American College of Cardiology*, *79*(4), 372–389.

Maron, B. J., & Maron, M. S. (2013). Hypertrophic cardiomyopathy. *The Lancet*, *381*(9862), 242–255.

Maron, B. J., McKenna, W. J., Danielson, G. K., Kappenberger, L. J., Kuhn, H. J., Seidman, C. E., et al. (2003). American College of Cardiology/European Society of Cardiology clinical expert consensus document on hypertrophic cardiomyopathy: a report of the American College of Cardiology foundation task force on clinical expert consensus documents and the European Society of Cardiology committee for practice guidelines. *Journal of the American College of Cardiology*, *42*(9), 1687–1713.

Maron, B. J., Rowin, E. J., Casey, S. A., & Maron, M. S. (2016). How hypertrophic cardiomyopathy became a contemporary treatable genetic disease with low mortality: shaped by 50 years of clinical research and practice. *JAMA Cardiology*, *1*(1), 98–105.

Menegotto, A. B., Becker, C. D. L., & Cazella, S. C. (2021). Computer-aided diagnosis of hepatocellular carcinoma fusing imaging and structured health data. *Health Information Science and Systems*, *9*(1), 1–11.

Morrow, A. G., & Brockenbrough, E. C. (1961). Surgical treatment of idiopathic hypertrophic subaortic stenosis: technic and hemodynamic results of subaortic ventriculotomy. *Annals of Surgery*, *154*(2), 181.

Mosqueira, D., Smith, J. G., Bhagwan, J. R., & Denning, C. (2019). Modeling hypertrophic cardiomyopathy: mechanistic insights and pharmacological intervention. *Trends in Molecular Medicine*, *25*(9), 775–790.

O'hara, R. P., Binka, E., Prakosa, A., Zimmerman, S. L., Cartoski, M. J., Abraham, M. R., et al. (2022). Personalized computational heart models with T1-mapped fibrotic remodeling predict sudden death risk in patients with hypertrophic cardiomyopathy. *Elife*, *11*, Article e73325.

Ouyang, D., Theurer, J., Stein, N. R., Hughes, J. W., Elias, P., He, B., et al. (2024). Electrocardiographic deep learning for predicting post-procedural mortality: a model development and validation study. *The Lancet Digital Health*, *6*(1), e70–e78.

Pandey, D., Wang, H., Yin, X., Wang, K., Zhang, Y., & Shen, J. (2022). Automatic breast lesion segmentation in phase preserved DCE-MRIs. *Health Information Science and Systems*, *10*(1), 1–19.

Park, S., & Chung, M. (2022). Cardiac segmentation on CT Images through shape-aware contour attentions. *Computers in Biology and Medicine*, *147*, Article 105782.

Payer, C., Štern, D., Bischof, H., & Urschler, M. (2017). Multi-label whole heart segmentation using CNNs and anatomical label configurations. In *International workshop on statistical atlases and computational models of the heart* (pp. 190–198). Springer.

Petitjean, C., Zuluaga, M. A., Bai, W., Dacher, J.-N., Grosgeorge, D., Caudron, J., et al. (2015). Right ventricle segmentation from cardiac MRI: a collation study. *Medical Image Analysis*, *19*(1), 187–202.

Pičulin, M., Smole, T., Žunkovič, B., Kokalj, E., Robnik-Šikonja, M., Kukar, M., et al. (2022). Disease progression of hypertrophic cardiomyopathy: modeling using machine learning. *JMIR Medical Informatics*, *10*, Article e30483.

Pollick, C., Shmueli, H., Maalouf, N., & Zadikany, R. H. (2020). Left ventricular cavity obliteration: mechanism of the intracavitary gradient and differentiation from hypertrophic obstructive cardiomyopathy. *Echocardiography*, *37*(6), 822–831.

Radau, P., Lu, Y., Connelly, K., Paul, G., Dick, A. J., & Wright, G. A. (2009). Evaluation framework for algorithms segmenting short axis cardiac MRI. *The MIDAS Journal*.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, *28*.

Ronneberger, O., Fischer, P., Brox, T., Brox, T., & Brox, T. (2015). *U-Net: Convolutional networks for biomedical image segmentation* (pp. 234–241). Springer International Publishing.

Rowin, E. J., Maron, B. J., Olivotto, I., & Maron, M. S. (2017). Role of exercise testing in hypertrophic cardiomyopathy. *JACC: Cardiovascular Imaging*, *10*(11), 1374–1386.

Schoendube, F. A., Klues, H. G., Reith, S., Flachskampf, F. A., Hanrath, P., & Messmer, B. J. (1995). Long-term clinical and echocardiographic follow-up after surgical correction of hypertrophic obstructive cardiomyopathy with extended myectomy and reconstruction of the subvalvular mitral apparatus. *Circulation*, *92*(9), 122–127.

Tuohy, C. V., Kaul, S., Song, H. K., Nazer, B., & Heitner, S. B. (2020). Hypertrophic cardiomyopathy: the future of treatment. *European Journal of Heart Failure*, *22*(2), 228–240.

Vainrib, A., Massera, D., Sherrid, M. V., Swistel, D. G., Bamira, D., Ibrahim, H., et al. (2021). Three-dimensional imaging and dynamic modeling of systolic anterior motion of the mitral valve. *Journal of the American Society of Echocardiography*, *34*(1), 89–96.

Veselka, J., Adla, T., Adlova, R., & Bartel, T. (2018). Three-dimensional heart printing for planning of septal reduction therapy in patients with hypertrophic obstructive cardiomyopathy. *International Journal of Angiology*, *27*(03), 165–166.

Wang, C., MacGillivray, T., Macnaught, G., Yang, G., & Newby, D. (2018). A two-stage 3D unet framework for multi-class segmentation on full resolution image. arXiv preprint arXiv:1804.04341.

Xu, Z., Wu, Z., & Feng, J. (2018). CFUN: Combining faster R-CNN and U-net network for efficient whole heart segmentation. arXiv preprint arXiv:1812.04914.

Xue, W., Brahm, G., Pandey, S., Leung, S., & Li, S. (2018). Full left ventricle quantification via deep multitask relationships learning. *Medical Image Analysis*, *43*, 54–65.

Yang, X., Bian, C., Yu, L., Ni, D., & Heng, P.-A. (2017). Hybrid loss guided convolutional networks for whole heart parsing. In *International workshop on statistical atlases and computational models of the heart* (pp. 215–223). Springer.

Ye, C., Wang, W., Zhang, S., & Wang, K. (2019). Multi-depth fusion network for whole-heart CT image segmentation. *IEEE Access*, *7*, 23421–23429.

Zhang, J., Jin, Y., Xu, J., Xu, X., & Zhang, Y. (2018). Mdu-net: Multi-scale densely connected u-net for biomedical image segmentation. arXiv preprint arXiv:1812.00352.

Zhang, J., Zhang, Y., & Xu, X. (2021). Pyramid U-net for retinal vessel segmentation. In *ICASSP 2021-2021 IEEE international conference on acoustics, speech and signal processing* (pp. 1125–1129). IEEE.

Zhou, H., Li, L., Liu, Z., Zhao, K., Chen, X., Lu, M., et al. (2021). Deep learning algorithm to improve hypertrophic cardiomyopathy mutation prediction using cardiac cine images. *European Radiology*, *31*(6), 3931–3940.

Zhuang, X., Li, L., Payer, C., Štern, D., Urschler, M., Heinrich, M. P., et al. (2019). Evaluation of algorithms for multi-modality whole heart segmentation: an open-access grand challenge. *Medical Image Analysis*, *58*, Article 101537.

Zhuang, X., & Shen, J. (2016). Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. *Medical Image Analysis*, *31*, 77–87.