Multi-View Resliced Image Fusion for Enhanced Cardiac MRI Segmentation

Shilajit Banerjee [1]
banerjee.shilajit@tcs.com
Divya Bhatia[2]
bhatia.divya@tcs.com
Vivek Chandel[3]
vivek.chandel@tcs.com
Aparna Kanakatte Gurumurthy[2]
aparna.kg@tcs.com
Aniruddha Sinha[1]
aniruddha.s@tcs.com
Sundeep Khandelwal [3]
sundeep.khandelwal@tcs.com
Avik Ghose [1]
avik.ghose@tcs.com

[1] Connected Digital Health
TCS Research
Kolkata India
[2] Connected Digital Health
TCS Research
Bangalore India
[3] Connected Digital Health
TCS Research
New Delhi India

Abstract

Cardiovascular diseases (CVD) are one of the leading causes of death worldwide, so early detection using advanced imaging techniques such as MRI is critical for administering effective treatment. However, segmentation of heart substructures remains challenging due to limited annotated data, image complexity, and variability across MRI scanners, protocols, patient positioning, and inter-radiologist variability. Manual annotation is time-consuming and prone to inconsistency. In this paper, we introduce a new 3D data augmentation method that creates extra slice views from 3D MRI scans, along with their corresponding ground truth masks. We fuse these augmented views with the original training data to enhance the dataset by providing multiple perspectives of heart substructures in medically relevant directions. This multi-view slice augmentation approach directly improves 2D and 3D segmentation performance by exposing the model to diverse anatomical viewpoints during training. The method improves model generalization across imaging variations and enhances segmentation accuracy. To support this claim, we test our method on a completely unseen dataset to test effectiveness of the augmentation. We demonstrate that our method outperforms standard augmentation techniques and generates more clinically relevant segmentation masks on completely unseen data. Thus it reduces manual segmentation time, helping radiologists handle new data more efficiently. Experiments on publicly available datasets confirm the effectiveness of our method with significant improvement in blind testing across multiple datasets.

Clinical Relevance: Accurate cardiac segmentation is essential for diagnosing and treating cardiovascular diseases. Our method enhances 2D segmentation by generating clinically

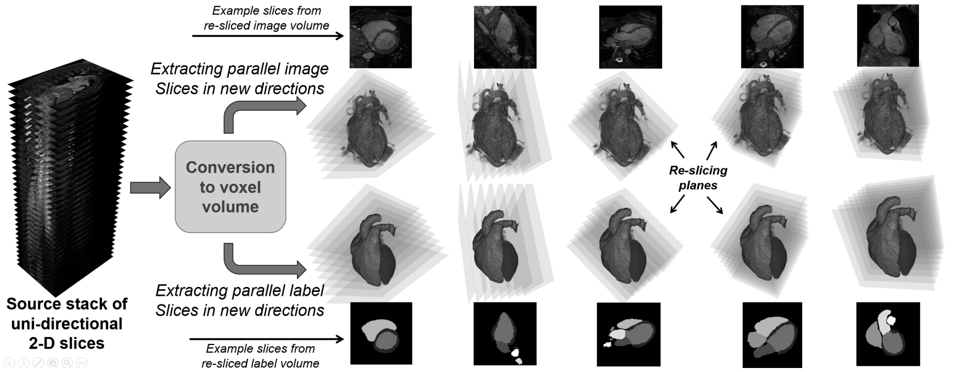

Figure 1: Generating stack of slices for images and labels in new directions from the original uni-directional stack. No. of slices in each stack shown here after voxel conversion do not reflect the actual number. New views from left to right are: 2-ch SAX, 2-ch LAX, 3-ch, 4-ch, LVOT.

relevant views like the 2-chamber short-axis (2-ch SAX), 2-chamber long-axis (2-ch LAX) etc. These views capture vital anatomical details needed to evaluate ventricular function, valvular integrity, and overall heart anatomy. Improved segmentation accuracy reduces the manual effort required by radiologists. This approach streamlines radiology workflows and supports digital twin modeling for personalized treatment. It also enables early and reliable detection of cardiac abnormalities. Ultimately, our method contributes to improved patient outcomes and more efficient clinical practices.

1 Introduction

Cardiovascular diseases (CVD) are one of the leading causes of morbidity and mortality worldwide, placing a significant burden on global healthcare systems. Timely diagnosis and accurate treatment planning are essential to reduce CVD-related complications and improve patient outcomes. One key aspect of personalized treatment is accurate segmentation of heart substructures, including Left Ventricle (LV), Right Ventricle (RV), Left Atrium (LA), Right Atrium (RA), Aorta, Myocardium (Myo), and Pulmonary Artery (PA). This process, known as whole-heart segmentation (WHS) from patient specific data, is crucial for generating personalized digital twin model of a patient's heart, which can assist pre-surgical planning, intervention planning, procedure simulation, and disease progression evaluation [7, 14, 19, 25, 33, 35].

However, accurate segmentation of heart substructures from medical imaging remains a significant challenge. A major limitation is scarcity of annotated data, as manual segmentation is both time-consuming and resource-intensive. High-quality, annotated datasets require expert involvement, and such datasets are often insufficient in size or diversity, limiting the generalizability of segmentation models. Furthermore, variability in medical imaging, including differences in MRI scanners, acquisition protocols, and patient positioning, can lead to inconsistencies in image quality, orientation, and resolution [7, 21, 32]. These variations are particularly challenging for deep learning (DL) models, which require large, diverse datasets to generalize effectively across different conditions. Studies such as [18] and [31] highlight how variability between different MRI machines and protocols can lead to inconsistent model performance and reduced segmentation accuracy.

To overcome the issue of data scarcity and variability, deep learning-based data augmentation techniques have been widely explored. These methods typically rely on generative models, such as GANs (Generative Adversarial Networks) or other sophisticated techniques, to synthesize new data samples from existing datasets [6, 8, 10, 15]. While these methods have shown promise in expanding datasets, they require extensive computational resources and time-consuming training processes. Additionally, the synthetic data generated by these models may not always capture the full range of realistic variations present in clinical datasets, limiting their utility in real-world applications. Common augmentation techniques, such as image rotation, flipping, scaling, zooming and intensity adjustments, are also employed to expand the dataset artificially. However, these methods fail to address more complex variations, such as differences in scan orientation or slice position, which are critical for accurate heart segmentation [23, 27, 34]. Recent works, such as those by [5] and [16], propose more sophisticated augmentation strategies that address these challenges, but they still rely on deep learning models, which may not always be feasible in resource-constrained settings.

In this work, we introduce an innovative approach to data augmentation that can be applied to most 3D structural data like CT, MRI, PET, etc. Our proposed augmentation technique helps in improving the accuracy of segmentation and classification models of DL. From an existing 3D MRI volume, we derive several medically relevant views that capture key anatomical structures of the heart, including its chambers, major arteries, and veins. We then fuse these newly generated images with the original dataset through concatenation, creating an enriched training set that provides comprehensive multi-view information for deep learning models. Specifically, we generate the following views: i) 2-chamber short-axis (2-ch SAX) view provides a cross-sectional image of the left ventricle. ii) 2-chamber long-axis (2-ch LAX) view extends the perspective to include the left atrium. iii) 3-chamber (3-ch) view incorporating the aortic valve, providing insight into both the myocardial and the valvular regions. iv) 4-chamber (4-ch) view simultaneously displaying all four heart chambers, which is essential to evaluate the overall size and function of the chamber and v) Left ventricular outflow tract (LVOT) view focusing on the left ventricular outflow tract and aortic valve to assess blood flow dynamics and detect potential obstructions. This enriched set of views provides the segmentation model with more comprehensive information about the structure of the heart. The view generation is shown in Fig 1, which is detailed in the methodology section. Notably, our method creates corresponding labeled images for these new slices and fuses them with the original dataset, ensuring that the augmented data integrates seamlessly with the original dataset. By leveraging the 3D context in a computationally efficient manner and fusing multiple view-generated images with original data, our approach significantly improves segmentation performance. This paper details our methodology and presents experimental results that highlight the advantages of using 3D-aware information fusion for cardiac MRI segmentation, towards improving the segmentation accuracy even with minimal dataset availability for training the DL networks. Major contributions and benefits of our work are summarized as follows:

- We propose a novel 3D-aware data augmentation method that generates new slice views from existing MRI data and fuses them with the original dataset, leveraging the underlying 3D anatomical context to address data scarcity and variability challenges in whole-heart segmentation.
- We introduce a method for generating corresponding ground truth labels for the newly generated slices and fusing this augmented data with the existing dataset, ensuring

seamless integration and maintaining accurate segmentation annotations.

- We show that our method significantly improves segmentation performance of heart substructures, outperforming traditional augmentation techniques such as flipping, rotation, and intensity adjustments for both 2D and 3D segmentation models.
- Through blind testing on other MRI data (not used in training), we illustrate that our method demonstrates model robustness, enabling effective generalization across different scanners, acquisition protocols, and patient positioning effectively helping the radiologists to save time on manual segmentation for new MRI scans.
- The proposed augmentation method can be extended to other 3D modalities such as CT and PET, and applied to different anatomical regions (e.g., brain, liver) to improve segmentation, classification, and overall predictive accuracy.

The rest of the paper is organized as follows. We present the details of the proposed method in Section 2. The details of the experiments and the results are given in Section 3. Finally, we conclude the paper in Section 4.

2 Methodology

This section presents the dataset characteristics, 3D-aware augmentation strategy, model architectures, and training procedures employed in this work. We detail each component to provide a complete framework for 3D-aware segmentation enhancement.

2.1 Dataset

We use the MM-WHS dataset [23] published at MICCAI 2017. The dataset consists of 20 anonymized clinical 3D MRI and 3D CT volumes of varying sizes. Each volume contains manual segmentation masks for structures including the right ventricle (RV), left atrium (LA), right atrium (RA), aorta, myocardium (Myo), and pulmonary artery (PA). The original volumes have different dimensions due to patient-specific anatomical variations. In the original dataset, the 3D MRI volumes consist of a total of 2,449 2D uni-directional slice images for all the subjects. To augment the dataset, we generate 5 different volumetric views for each patient volume, resulting in a total of 9,416 multi-directional 2D slice images. Owing to different slice thickness values in the original slices and generated ones, the latter no. of slices are less than the expected 5 times the former. We employ an 80-20 patient-level split for training and validation to ensure no data leakage between sets. We apply 2D segmentation techniques on each of these slices individually. To ensure repeatability, we employ 5-fold cross-validation in our experiments. Additionally, we perform blind testing on multiple datasets. These include the Sunnybrook dataset [24], which was released as part of an LV segmentation challenge. We also perform blind testing on ACDC [1] and M&Ms [3]. These datasets contain three labels: LV, RV and Myo. We use these datasets to evaluate our method's ability to generalize to unseen datasets, establishing its utility in practice.

2.2 Proposed 3D Aware Multiview Data Augmentation

For every subject in the MM-WHS dataset, there is a collection of 2D slice images representing a 3-D volume stacked in a single slicing direction (as detailed in section 2.1 above).

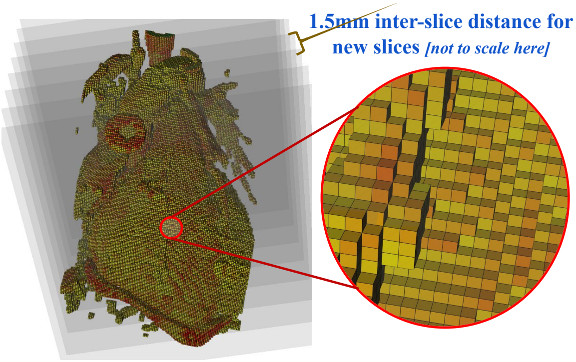

Figure 2: Left: Heart volume of a subject visualized with voxels. Right: Magnified view of voxel collection. Each voxel cube is 1x1x1mm in dimension.

It is noted that this limits the models capability to establish spatial correlations in directions other than the pre-set slicing direction.

We propose to re-slice the existing stack of 2D slices in multiple directions to provide new augmenting stacks, encompassing an entirely new spatial context. This is similar to rotational augmentation of 3D volume instances for segmentation models which receive 3D input. But, our modeling approach aims at 2D image slice as an input instead of a 3D volume. Hence, re-slicing becomes a requisite. In addition, the augmenting slicing directions are not chosen randomly, but along certain medically relevant well-known viewing directions, as enlisted in section 1. These views together provide a coherent context of heart's anatomy, assisting cardiac experts in critical assessments relating to ejection fraction and morphometry. Figure 1 depicts the original slices present in the dataset, followed by volumes generated from image and label slices in the dataset. Also depicted are sets of parallel re-slicing planes in grey color used to generate stacks of new slices in order to augment with richer spatial context. Let the collective set of pixels of all the stacked 2D slices in a volume be represented as $\mathbb{P}$. 3D volumes in the dataset are stored in Nifti (.nii) format, which includes affine transformation matrix, M_{aff} along with the pixel data. Every pixel $p \in \mathbb{P}$ with coordinate $(i, j, k) \in \mathbb{I}^3$ is transformed to RAS (right, anterior, superior) system with coordinate $(x_p, y_p, z_p) \in \mathbb{R}^3$ with respect to the physical location of the imaging machine as:

$$\begin{bmatrix} x_p & y_p & z_p & 1 \end{bmatrix}^T = M_{aff} \cdot \begin{bmatrix} i & j & k & 1 \end{bmatrix}^T \tag{1}$$

After spatial transformation of the stacked slices, the represented 3D volume is then converted into a voxel image volume with the same bounding area as the stack before re-slicing for a coherent distribution of raw pixel values eliminating the effect of varied spatial resolution in different axes. We use voxel dimension of $1 \times 1 \times 1mm$ to avoid any loss of spatial information (Fig 2 depicts individual voxels with heart volume visualized in voxel space), given that the inter-slice and inter-pixel resolution in the dataset ranges between $1mm - 1.6mm$. We use inverse distance weighing interpolation [26] technique for estimating the value for a voxel from the pixel values in the transformed stack. Let the centroid's location and value of i^{th} voxel be $\mathbf{x_i}, a_i$, and location and value of j^{th} pixel in $\mathbb{P}$ be represented as $\mathbf{y_j}, b_j$. Voxel value a_i is then calculated as:

$$a_i = \frac{\sum_{j=1}^{N} w_j \cdot b_j}{\sum_{j=1}^{N} w_j} \tag{2}$$

$$\text{where, } w_j = \left(\frac{1}{dist(\mathbf{x_i}, \mathbf{y_j})}\right)^p \text{ and,}$$

$\{j, j+1, j+2, ...N\}$ are the set of N pixels in $\mathbb{P}$ located closest to $\mathbf{x_i}$. For cardiac images in the dataset, we use $p = 2$ and $N = 2$ in order to capture the pixel inputs within a $1mm^3$ voxel space. Similar steps are followed for the voxel label volumes, but with $N = 1$ to avoid contamination of the actual label values due to multi-pixel interpolation.

The voxel volume is now sliced from five different views in contrast to the slices in original dataset, which are present in only one of the body axes. The newly projected slices are created at an inter-slice distance of $1.5mm$ as depicted in Fig 2.

In order to perform the slicing in a new direction using a slicing plane, a coordinate axes frame, $\mathcal{F}_n$ is calculated with respect to the slicing plane. This is executed such that one of the basis planes of $\mathcal{F}_n$ (e.g. XZ) coincides with the slicing plane. Let $\mathcal{F}_v$ be the coordinate frame of the volume to be sliced. Now, voxel locations of the volume are transformed from $\mathcal{F}_v$ to $\mathcal{F}_n$. The value for every pixel in the new slice is then calculated by interpolating voxels in the volume which are nearest to the pixel. These voxels are estimated using intersection of the transformed volume with the chosen basis plane. The same process is carried out for the ground truth label volumes as well, but instead of interpolating the nearby voxel values, the value of the nearest voxel is chosen, which is critical in order to preserve the original label values in the dataset.

2.3 Models and Training

The proposed augmentation is evaluated using DeepLabV3+ [4] with a ResNet-152 [9] backbone, pre-trained on ImageNet, nnU-Net [11] and earlier proposed 3D GAN [13] for substructure segmentation.

2.3.1 Training of DeepLabV3+, nnU-Net and 3D GAN

For heart substructure segmentation, we employ DeepLabV3+ [4] with ResNet-152 [9] backbone and nnU-Net [11]. DeepLabV3+ enhances feature extraction through deep residual learning while mitigating vanishing gradients. The encoder-decoder structure captures rich hierarchical features while refining segmentation maps through atrous convolutions. nnU-Net provides a self-configuring framework that automatically adapts preprocessing, network architecture, and training procedures to dataset characteristics. Recent advances demonstrate the potential of custom 3D GAN architectures for medical image segmentation tasks [13]. We adopt a 3D GAN framework for volumetric heart substructure segmentation.

Training utilizes cross-entropy loss for multi-class segmentation in DeepLabV3+ and nnU-Net:

$$\mathcal{L} = -\sum_{c=1}^{C} y_c \log(p_c) \tag{3}$$

where C represents class (Left Ventricle, Right Ventricle, etc.), y_c denotes ground truth, and p_c indicates predicted probability for class c.

Table 1: Results on MM-WHS [23] dataset for segmentation using three models: DeepLabV3+ with ResNet151 as the backbone, nnU-Net, and 3D GAN. The evaluation is based on the Dice Score, reported as the mean and standard deviation from 5-fold cross-validation.

Method/Model	**Myo**	**LA**	**LV**	**RA**	**RV**	**Aorta**	**PA**	**Average**
DeepLabV3+ without Augmentation	0.855± 0.019	0.771± 0.071	0.845± 0.036	0.888± 0.067	0.920± 0.033	0.887± 0.016	0.886± 0.025	0.865± 0.016
DeepLabV3+ with Standard Augmentation	0.862± 0.012	0.778± 0.062	0.852± 0.045	0.895± 0.054	0.927± 0.025	0.895± 0.021	0.894± 0.032	0.873± 0.063
DeepLabV3+ with Proposed Augmentation	0.940± 0.009	0.902± 0.014	0.923± 0.011	0.927± 0.006	0.947± 0.008	0.897± 0.025	0.893± 0.018	0.911± 0.012
nnU-Net without Augmentation	0.845± 0.044	0.930± 0.012	0.905± 0.018	0.875± 0.036	0.755± 0.057	0.853± 0.090	0.865± 0.073	0.861± 0.043
nnU-Net with Standard Augmentation	0.863± 0.078	0.914± 0.022	**0.963**± 0.012	0.895± 0.019	**0.935**± 0.006	0.931± 0.010	0.770± 0.068	0.896± 0.031
nnU-Net with Proposed Augmentation	0.868± 0.016	0.881± 0.024	0.953± 0.010	**0.920**± 0.030	0.933± 0.013	0.938± 0.009	**0.927**± 0.026	0.917± 0.017
3D GAN without Augmentation	0.730± 0.190	0.750± 0.050	0.860± 0.320	0.820± 0.130	0.840± 0.090	0.870± 0.110	0.790± 0.220	0.809± 0.160
3D GAN with Standard Augmentation	0.820± 0.230	0.800± 0.140	0.850± 0.070	0.770± 0.440	0.830± 0.280	0.810± 0.570	0.800± 0.150	0.811± 0.270
3D GAN with Proposed Augmentation	**0.943**± 0.060	**0.951**± 0.120	0.960± 0.210	0.933± 0.080	0.925± 0.140	**0.962**± 0.160	0.908± 0.240	**0.940**± 0.150

Table 2: Comparison of segmentation performance on the MM-WHS [23] dataset with prior methods. The table presents Dice score results from both 2D and 3D segmentation models.

Method/Model	**Myo**	**LA**	**LV**	**RA**	**RV**	**Aorta**	**PA**	**Average**
Chunliang *et al.* (2D) [illegible]	0.807± 0.059	0.847± 0.061	0.895± 0.057	0.821± 0.087	0.795± 0.102	0.679± 0.180	0.743± 0.204	0.798
Mortazi *et al.* (2D) [illegible]	0.825	0.887	0.932	0.874	0.884	0.772	0.784	0.850
Wang *et al.* (3D) [illegible]	0.744	0.852	0.863	0.840	0.849	0.824	0.788	0.832
Payer *et al.* (3D) [illegible]	0.778	0.855	0.916	0.881	0.868	0.888	0.731	0.863
Aparna *et al.* (3D) [illegible]	0.839	0.855	0.938	0.877	0.914	0.824	0.797	0.881
DeepLabV3+ with Proposed Augmentation (2D)	**0.940± 0.009**	0.902± 0.014	0.923± 0.011	0.927± 0.006	0.947± 0.008	0.897± 0.025	0.893± 0.018	0.918± 0.013
nnU-Net with Proposed Augmentation	0.868± 0.016	0.881± 0.024	0.953± 0.010	0.920± 0.030	0.933± 0.013	0.938± 0.009	**0.927± 0.026**	0.917± 0.018
3D GAN with Proposed Augmentation (3D)	0.943± 0.062	**0.951± 0.123**	**0.960± 0.214**	**0.933± 0.085**	**0.925± 0.142**	**0.962± 0.168**	0.908± 0.245	**0.940± 0.148**

For 3D GAN training, the discriminator loss $\mathcal{L}_D$ is defined as:

$$\mathcal{L}_D = \frac{1}{2}\Big[\mathrm{MSE}\big(D(I, y_{\mathrm{gt}}), 1\big) + \mathrm{MSE}\big(D(I, y_{\mathrm{pred}}), 0\big)\Big] \tag{4}$$

The discriminator learns to distinguish real segmentation from generated one by assigning scores close to 1 for real pairs and 0 for fake pairs. The generator loss $\mathcal{L}_G$ combines adversarial and segmentation terms:

$$\mathcal{L}_G = \mathrm{MSE}\big(D(I, y_{\mathrm{pred}}), 1\big) + \alpha \cdot \mathrm{GDL}(y_{\mathrm{gt}}, y_{\mathrm{pred}}) \tag{5}$$

The adversarial term encourages the generator to produce realistic segmentations while the Generalized Dice Loss (GDL) [28] term ensures spatial accuracy. Here I represents input MRI image, y_{gt} denotes ground truth segmentation, y_{pred} indicates generator prediction, and α balances adversarial and segmentation loss contributions.

3 Results and Discussion

3.1 Implementation Details

We use PyTorch for training and optimization with Adam optimizer [17]. For DeepLabV3+, parameters include $\beta_1 = 0.9$, $\beta_2 = 0.999$, learning rate 2×10^{-4}. We train it for 300 epochs with batch size 16. For nnU-Net, we use $\beta_1 = 0.9$, $\beta_2 = 0.999$, learning rate 3×10^{-4}. We train it for 200 epochs with batch size 8. The 3D-GAN uses learning rate 2×10^{-4} with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\alpha = 4$. We take batch size 8 and train it for 400 epochs. A dropout rate of 0.4 mitigates mode collapse risk. All models undergo five-fold cross-validation with training on NVIDIA V100 GPU. Each subset serves once for testing with training performed on the remaining four subsets.

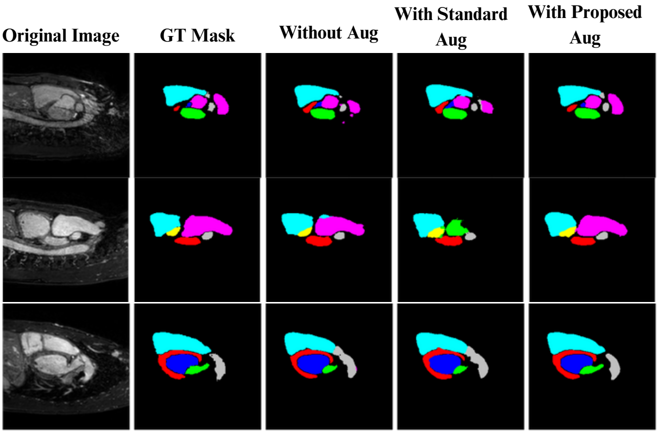

Figure 3: Comparison of Ground Truth and Predicted Masks for MM-WHS [23] Dataset from two models: DeepLabV3+ (First row), nnU-Net (2nd Row) and 3D GAN (3rd row). The first column shows the original image, the second column displays the ground truth (GT) mask, the third column presents the predicted mask without augmentation, the fourth column shows the predicted mask with standard augmentation, and the fifth column shows the predicted mask using the proposed augmentation method.

3.2 Performance Evaluation

To show the impact of our proposed data augmentation, we compare performance of the three DL models with three different input data modalities: 1) no augmentation, 2) standard augmentation and 3) proposed augmentation.

3.2.1 Comparison with Standard Augmentation

Standard augmentation techniques include random horizontal and vertical flips (50% probability), rotations (±30°), and brightness adjustments (scaling factors 0.8-1.2). Performance is evaluated using Dice score which quantifies overlap between predicted segmentation and ground truth. Table 1 presents the results. DeepLabV3+ with standard augmentation modestly improves the average Dice score from 0.865 to 0.873 when compared with no augmentation. Proposed augmentation yields significantly higher performance with DeepLabV3+ achieving an average Dice score of 0.911. nnU-Net shows improvement in average Dice score from 0.861 without augmentation to 0.896 with standard augmentation and reaches 0.917 with proposed augmentation. 3D GAN experiences performance variation with standard augmentation (0.811 compared to 0.809 baseline). However, proposed augmentation yields substantially higher scores with 3D GAN achieving an average Dice score of 0.940. Table 1 clearly shows the impact of our proposed clinically relevant augmentation with a good improvement in the average Dice score. The new augmentation scheme helped in better generalization of the model owing to better captured spatial context in the data than the normally used standard augmentation.

3.2.2 Comparison with Prior Methods on MM-WHS

We compare our segmentation performance with proposed augmentation against state-of-the-art approaches on the MM-WHS dataset [23]. Table 2 presents Dice scores for cardiac

Table 3: Cross-dataset generalization performance: Models trained on MM-WHS [23] and tested on multiple datasets.

Method/Model	**Sunnybrook**				**ACDC**				**M&M**			
	LV	**RV**	**Myo**	**Avg**	**LV**	**RV**	**Myo**	**Avg**	**LV**	**RV**	**Myo**	**Avg**
DeepLabV3+ w/o Aug	0.387	-	-	0.387	0.631	0.436	0.223	0.430	0.250	0.230	0.083	0.187
DeepLabV3+ w/ Std Aug	0.472	-	-	0.472	0.763	0.625	0.678	0.689	0.428	0.409	0.285	0.374
DeepLabV3+ w/ Prop Aug	**0.806**	-	-	**0.806**	**0.891**	**0.751**	**0.702**	**0.781**	**0.667**	**0.529**	**0.583**	**0.593**
nnU-Net w/o Aug	0.614	-	-	0.614	0.553	0.327	0.486	0.455	0.442	0.205	0.053	0.233
nnU-Net w/ Std Aug	0.773	-	-	0.773	0.752	0.608	0.665	0.675	0.516	0.354	0.283	0.384
nnU-Net w/ Prop Aug	**0.885**	-	-	**0.885**	**0.843**	**0.576**	**0.754**	**0.724**	**0.725**	**0.573**	**0.457**	**0.585**
3D GAN w/o Aug	0.574	-	-	0.574	0.546	0.094	0.305	0.315	0.527	0.083	0.224	0.278
3D GAN w/ Std Aug	0.706	-	-	0.706	0.715	0.446	0.623	0.595	0.634	0.285	0.353	0.424
3D GAN w/ Prop Aug	**0.836**	-	-	**0.836**	**0.784**	**0.587**	**0.653**	**0.675**	**0.763**	**0.556**	**0.496**	**0.605**

structures across 2D and 3D segmentation models. Our approach enhances overall performance while improving class-wise accuracy, particularly for left and right ventricles. 3D GAN with proposed augmentation achieves the highest overall Dice score of 0.940. nnU-Net with proposed augmentation reaches an average Dice score of 0.920. DeepLabV3+ with proposed augmentation attains an average Dice score of 0.911. All three methods outperform existing state-of-the-art approaches significantly. Proposed augmentation reduces inter-class variability, ensuring uniform segmentation quality across cardiac structures. Prior methods show higher inter-class variability and yield less consistent outcomes. A more homogeneous performance from our approach indicates superior robustness in capturing anatomical variations. Our models maintain low class-wise variability, reinforcing consistent anatomical detail capture. This consistency is critical for reliable diagnosis and treatment planning.

3.2.3 Performance in Blind Testing

Blind testing evaluates cross-dataset generalization on Sunnybrook [24], ACDC [1], and M&M [3] datasets. Models are trained on MM-WHS dataset [23] containing only 20 volumes. Since Sunnybrook contains 2D data, we stack the slices to create 3D volumes for 3D GAN evaluation. The M&M dataset presents particularly challenging segmentation scenarios due to data acquisition from different scanners and imaging protocols. Table 3 presents comprehensive results comparing standard against proposed augmentation approaches across three methods and three testing datasets. On the Sunnybrook dataset, our approach demonstrates substantial improvements over standard augmentation. DeepLabV3+ achieves an average Dice score of 0.806 versus 0.472 (70.8% improvement). nnU-Net reaches 0.885 compared to 0.773 (14.5% improvement). 3D GAN attains 0.836 versus 0.706 (18.4% improvement). ACDC dataset results demonstrate significant performance gains with our method. DeepLabV3+ average Dice improves from 0.689 to 0.781 (13.3% improvement). nnU-Net advances from 0.675 to 0.724 (7.3% improvement). 3D GAN increases from 0.595 to 0.675 (13.4% improvement). The M&M dataset also shows notable enhancements. DeepLabV3+ average Dice rises from 0.374 to 0.593 (58.6% improvement). nnU-Net improves from 0.384 to 0.585 (52.3% improvement). 3D GAN advances from 0.424 to 0.605 (42.7% improvement).These substantial improvements achieved using only 20 training volumes highlight the effectiveness of our proposed augmentation approach compared to standard augmentation techniques. The approach reduces radiologists' manual segmentation time and streamlines clinical workflows.

4 Conclusion and Future Work

This paper introduced novel 3D-aware data augmentation enhancing heart substructure segmentation from MRI scans. Generating additional slice views from existing 3D data improved training datasets and captured vital anatomical context. Experimental results demonstrated out performance over standard augmentation techniques, yielding more accurate segmentation and reducing radiologists' manual correction time. The multi-view augmentation technique can be adapted readily to other 3D imaging modalities (CT, PET) and anatomical regions (brain, liver) to enhance segmentation, classification, and related tasks. Future work explores integration with various deep learning architectures and application to different medical imaging modalities. Development of user-friendly clinical tools aims to further improve diagnostic accuracy and cardiovascular healthcare outcomes.

References

[1] Olivier Bernard, Alain Lalande, Clement Zotti, Frederick Cervenansky, Xin Yang, Pheng-Ann Heng, Irem Cetin, Karim Lekadir, Oscar Camara, Miguel Angel Gonzalez Ballester, et al. 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, 2018.

[2] N. Byrne, M. Velasco Forte, A. Tandon, I. Valverde, and T. Hussain. A systematic review of image segmentation methodology, used in the additive manufacture of patient-specific 3d printed models of the cardiovascular system. In *JRSM cardiovascular disease*, volume 5, page 2048004016645467, 2016.

[3] M Campello and K Lekadir. Multi-centre multi-vendor & multi-disease cardiac image segmentation challenge (m&ms). In *Medical Image Computing and Computer Assisted Intervention*, 2020.

[4] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *arXiv preprint arXiv:1802.02611*, 2018.

[5] P. Chlap, H. Min, N. Vandenberg, J. Dowling, L. Holloway, and A. Haworth. A review of medical image data augmentation techniques for deep learning applications. In *Journal of Medical Imaging and Radiation Oncology*, volume 65, pages 545–563, 2021.

[6] V. Fernandez, W.H.L. Pinaya, P. Borges, P.D. Tudosiu, M.S. Graham, T. Vercauteren, and M.J. Cardoso. Can segmentation models be trained with fully synthetically generated data? In *International Workshop on Simulation and Synthesis in Medical Imaging*, pages 79–90, Cham, 2022. Springer International Publishing.

[7] G. Foti and C. Longo. Deep learning and ai in reducing magnetic resonance imaging scanning time: advantages and pitfalls in clinical practice. In *Polish Journal of Radiology*, volume 89, page e443, 2024.

[8] F. Garcea, A. Serra, F. Lamberti, and L. Morra. Data augmentation for medical imaging: A systematic literature review. In *Computers in Biology and Medicine*, volume 152, page 106391, 2023.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CoRR*, volume abs/1512.03385, 2015. arXiv preprint arXiv:1512.03385.

[10] Z. Huang, Q. Yang, M. Tian, and Y. Gao. Synthesizing images with annotations for medical image segmentation using diffusion probabilistic model. In *International Journal of Imaging Systems and Technology*, volume 35, page e70007, 2025.

[11] Fabian Isensee, Jens Petersen, Andre Klein, David Zimmerer, Paul F Jaeger, Simon Kohl, Jakob Wasserthal, Gregor Koehler, Tobias Norajitra, Sebastian Wirkert, et al. nnu-net: Self-adapting framework for u-net-based medical image segmentation. *arXiv preprint arXiv:1809.10486*, 2018.

[12] A. Kanakatte, D. Bhatia, and A. Ghose. Heart region segmentation using dense vnet from multimodality images. In *43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3255–3258, Mexico, 2021.

[13] A. Kanakatte, D. Bhatia, and A. Ghose. 3d cardiac substructures segmentation from cmri using generative adversarial network (gan). In *44th Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, pages 1698–1701, Glasgow, Scotland, United Kingdom, 2022.

[14] D. Kang, J. Woo, P.J. Slomka, D. Dey, G. Germano, and C.C. Jay Kuo. Heart chambers and whole heart segmentation techniques. In *Journal of Electronic Imaging*, volume 21, pages 010901–010901, 2012.

[15] A. Kebaili, J. Lapuyade-Lahorgue, and S. Ruan. Deep learning approaches for data augmentation in medical imaging: a review. In *Journal of Imaging*, volume 9, page 81, 2023.

[16] A. Kebaili, J. Lapuyade-Lahorgue, and S. Ruan. Deep learning approaches for data augmentation in medical imaging: a review. In *Journal of Imaging*, volume 9, page 81, 2023.

[17] D.P. Kingma. Adam: A method for stochastic optimization. In *arXiv preprint arXiv:1412.6980*, 2014.

[18] G. Litjens, F. Ciompi, J.M. Wolterink, B.D. de Vos, T. Leiner, J. Teuwen, and I. Išgum. State-of-the-art deep learning in cardiovascular image analysis. In *JACC: Cardiovascular imaging*, volume 12, pages 1549–1565, 2019.

[19] Oishee Mazumder, Ayan Mukherjee, Shilajit Banerjee, Sundeep Khandelwal, K. M. Mandana, and Aniruddha Sinha. Personalization of a hemodynamic cardiac digital twin: An echocardiogram based approach. In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 5523–5530, 2024. doi: 10.1109/BIBM62325.2024.10822244.

[20] A. Mortazi, J. Burt, and U. Bagci. Multi-planar deep segmentation networks for cardiac substructures from mri and ct. In *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges: 8th International Workshop, STACOM 2017*, pages 199–206. Springer International Publishing, 2018.

[21] A.M. Owrangi, P.B. Greer, and C.K. Glide-Hurst. Mri-only treatment planning: benefits and challenges. In *Physics in Medicine and Biology*, volume 63, page 05TR01, 2018.

[22] C. Payer, D. Stern, H. Bischof, and M. Urschler. Multi-label whole heart segmentation using cnns and anatomical label configurations. In M. Pop, editor, *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges. STACOM 2017*, volume 10663 of *Lecture Notes in Computer Science*, pages 190–198. Springer, 2018.

[23] M. Pop, M. Sermesant, P.M. Jodoin, A. Lalande, X. Zhuang, G. Yang, A. Young, and O. Bernard. Statistical atlases and computational models of the heart. acdc and mmwhs challenges: 8th international workshop, stacom 2017. In *Proceedings of the 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017*, volume 10663, Quebec City, Canada, 2018. Springer.

[24] P. Radau, Y. Lu, K. Connelly, G. Paul, A.J. Dick, and G.A. Wright. Evaluation framework for algorithms segmenting short axis cardiac mri. In *The MIDAS Journal – Cardiac MR Left Ventricle Segmentation Challenge*. http://hdl.handle.net/10380/3070.

[25] A. Rouhollahi, J.N. Willi, S. Haltmeier, A. Mehrtash, R. Straughan, H. Javadikasgari, J. Brown, A. Itoh, K.I. de la Cruz, E. Aikawa, and E.R. Edelman. Cardiovision: a fully automated deep learning package for medical image segmentation and reconstruction generating digital twins for patients with aortic stenosis. In *Computerized Medical Imaging and Graphics*, volume 109, page 102289, 2023.

[26] Donald Shepard. A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 23rd ACM national conference*, pages 517–524, 1968.

[27] C. Shorten and T.M. Khoshgoftaar. A survey on image data augmentation for deep learning. In *Journal of big data*, volume 6, pages 1–48, 2019.

[28] Carole Helene Sudre et al. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep learning in medical image analysis and multimodal learning for clinical decision support : Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, held in conjunction with MICCAI 2017*, pages 240–248, Quebec City, QC, 2017.

[29] C. Wang and Ö. Smedby. Automatic whole heart segmentation using deep learning and shape context. In *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges. STACOM 2017*, volume 10663 of *Lecture Notes in Computer Science*, Cham, 2018. Springer.

[30] C. Wang, T.J. MacGillivray, G. Macnaught, G. Yang, and D.E. Newby. A two-stage 3d unet framework for multi-class segmentation on full resolution image. In *ArXiv*, volume abs/1804.04341, 2018.

[31] J. Wang, H. Zhu, S.H. Wang, and Y.D. Zhang. A review of deep learning on medical image analysis. In *Mobile Networks and Applications*, volume 26, pages 351–380, 2021.

[32] M. Zaitsev, J. Maclaren, and M. Herbst. Motion artifacts in mri: A complex problem with many partial solutions. In *Journal of Magnetic Resonance Imaging*, volume 42, pages 887–901, 2015.

[33] R. Zeleznik, J. Weiss, J. Taron, C. Guthier, D.S. Bitterman, C. Hancox, B.H. Kann, D.W. Kim, R.S. Punglia, J. Bredfeldt, and B. Foldyna. Deep-learning system to improve the quality and efficiency of volumetric heart segmentation for breast cancer. In *NPJ digital medicine*, volume 4, page 43, 2021.

[34] X. Zhuang, L. Li, C. Payer, D. Štern, M. Urschler, M.P. Heinrich, J. Oster, C. Wang, Ö. Smedby, C. Bian, and X. Yang. Evaluation of algorithms for multi-modality whole heart segmentation: an open-access grand challenge. In *Medical image analysis*, volume 58, page 101537, 2019.

[35] X. Zhuang, L. Li, C. Payer, D. Štern, M. Urschler, M.P. Heinrich, J. Oster, C. Wang, Ö. Smedby, C. Bian, X. Yang, P.A. Heng, A. Mortazi, U. Bagci, G. Yang, C. Sun, G. Galisot, J.Y. Ramel, T. Brouard, Q. Tong, W. Si, X. Liao, G. Zeng, Z. Shi, G. Zheng, C. Wang, T. MacGillivray, D. Newby, K. Rhode, S. Ourselin, R. Mohiaddin, J. Keegan, D. Firmin, and G. Yang. Evaluation of algorithms for multi-modality whole heart segmentation: An open-access grand challenge. In *Medical Image Analysis*, volume 58, page 101537, 2019.