SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V11I9P118

Volume 11 Issue 9, 207-214, September 2024

Original Article

U-Net Based Deep Learning Approach for 2D Cardiovascular Image Segmentation

P. Sudheer[1], K. Indumathy[2], Pallapati Ravi Kumar[3], J. Manoranjini[4], Vijayalaxmi Bindla[5], Birjis.Fathima[6]

[1]*Department of CSE (AI&ML), CVR College of Engineering Hyderabad, Telangana, India.*
[2]*Department of Computer Applications Christ College of Engineering & Technology, Chennai.*
[3]*Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Andhra Pradesh, India.*
[4]*Department of Artificial Intelligence and Data Science, Rajalakshmi Engineering College, Chennai.*
[5]*Department of CSE (AI&ML), G Narayanamma Institute of Technology, Hyderabad, Telangana, India.*
[6]*Department of CSE (Cyber Security), Swami Vivekananda Institute of Technology Telangana, India.*

[1]*Corresponding Author : sudheerchanty7@gmail.com*

Received: 08 July 2024 Revised: 18 August 2024 Accepted: 09 September 2024 Published: 30 September 2024

Abstract *- The goal of medical image segmentation is to organize pixels into several areas according to the several characteristics of the images. Due to several factors, such as variations in data signal-to-noise ratios, signal intensities, and individual variations in heart morphologies, segmenting 2D echo cardiovascular images remains a difficult process. This research introduces 3D U-Net-SparseVoxNet, a unique and effective 3D sparse convolutional network based on U-Net. Any two layers in this network that have the same feature map size may have direct connections with each other, and there are fewer connections overall. Consequently, by drastically reducing the network depth and ultimately utilizing a spatial self-attention mechanism to improve feature representation, 3D U-Net-SparseVoxNet can successfully handle the optimization issue of gradients disappearing when using a limited sample of data to train a 3D deep neural network architecture. This research presents a detailed evaluation of the suggested technique using the HVSMR 2016 dataset. The strategy performs better when compared to other approaches. The proposed method proved to provide good and efficient results in classifying the data with an accuracy of 90% compared to 3D U-Net and VoxResNet, with 74% and 80% accuracy, respectively.*

Keywords *- 2-D Echocardiography, U-Net, Segmentation, Images, Convolution neural network.*

1. Introduction

Multi-objective segmentation is a crucial technique in the fields of clinical applications, medical image processing, and disease diagnosis. Its primary goal is to segment images into distinct regions based on shared features or specific characteristics such as edges, structures, or shapes. Accurate segmentation is vital for precise diagnosis, effective prognostic predictions, and enhanced surgical planning. In recent years, deep learning models, particularly Convolution Neural Networks (CNNs), have been employed to tackle image segmentation challenges [4]. These models have demonstrated significant advancements in medical imaging, yet they encounter substantial limitations that impede their effectiveness in real-world scenarios.

One major issue is the challenge posed by insufficient edge information and ambiguous boundaries, which hinder the accurate delineation of structures in medical images. Additionally, low-quality images further exacerbate the difficulty of achieving reliable segmentation. Although architectures such as VoxResNet and U-Net have made strides in addressing these challenges, they still face significant hurdles. VoxResNet, despite its high accuracy and efficiency, is hampered by its deep architecture, which requires extensive computational resources and memory, and its lengthy training times [5]. U-Net, on the other hand, suffers from issues related to redundant information and inefficiencies in preserving pixel-level details, which limits its effectiveness in handling complex image segmentation tasks [6].

Addressing these gaps, this study introduces a novel approach through the development of 3D U-Net-3D U-Net-SparseVoxNet, an advanced 3D sparse convolution network designed to enhance medical image segmentation. Unlike traditional models, our proposed approach incorporates sparse convolution techniques along with self-attention mechanisms. This integration not only improves feature map representation but also captures long-range dependencies more effectively. By reducing the model parameters and eliminating unnecessary processing, our approach aims to achieve higher segmentation accuracy, faster convergence, and reduced risk of overfitting. This represents a significant advancement over existing methods, offering a more efficient and accurate solution for complex medical image segmentation tasks.

The paper is organized as follows: Section 1 outlines the proposed work and identifies the research gap. Section 2 provides a comprehensive review of related literature. Section 3 details the proposed methods, including the sparse block architecture and model design. Section 4 presents the experimental setup and findings, and Section 5 concludes the study with a summary of results and future directions.

2. Related Works

In the realm of multi-objective image segmentation, various methods have been explored to improve accuracy and efficiency. Early approaches, such as those proposed by Pham et al. [9], employed region-based active contour algorithms combined with fuzzy entropy clustering to segment brain images. Hongwei et al. [10] introduced a multi-objective clustering method coupled with a toroidal model-guided tracking approach to distinguish intricate features in vascular structures. These early efforts laid the groundwork for more advanced techniques, but challenges remained in handling complex and noisy data.

The advent of deep learning has revolutionized medical image segmentation, with significant contributions from various researchers. Çiçek et al. [11] utilized a dual-network approach to segment cardiac images, improving performance by localizing and distinguishing substructures within the heart. Ding et al. [13] incorporated attention mechanisms into their models to enhance classification accuracy while reducing computational costs. Despite these advancements, issues such as redundant information and limited effectiveness in noisy conditions persisted, highlighting the need for more refined solutions.

Recent developments have focused on automated image processing and quality evaluation. For instance, [18] explored anisotropic diffused filters for automated processing, while [19] evaluated various de-speckling filters for carotid plaque ultrasound images. These studies contributed valuable insights into image quality improvement but did not fully address the complexities of medical image segmentation.

Our study builds upon these advancements by introducing a customized light multi-head model designed for echo-specific representation and real-time processing. This model outperforms previous methods by providing greater accuracy and efficiency in segmenting complex medical images. By addressing the limitations of existing techniques and incorporating innovative features, our approach offers a more robust solution for tackling the challenges associated with medical image segmentation.

3. Proposed Methodologies

3.1. The Architecture of 3D U-Net-SparseVoxNet

The 3D U-Net-SparseVoxNet design that is suggested in this article. It enhances U-Net, which uses procedures for performing end-to-end training using up sampling and down sampling. Since when the characteristic maps vary in size, the sparse block loses all significance, padding is utilized to keep feature-map sizes.

Constant throughout all sparse blocks. Dilated convolutions are thus utilized in the last three layers, and conventional convolutions in the first four levels of each sparse block. The most significant components of the original feature map are highlighted using a spatial feature map.

Compared to conventional convolution, dilated convolution contains an additional hyperparameter that provides holes to the traditional convolution kernel. In this study, we add 3D data to the dilated convolution algorithm by combining it with regular convolution. We use three levels of dilation convolution with two, three, and five holes in addition to four layers of 3 × 3 × 3 conventional convolutions concerning the Dense VoxNet. The three stages of dilated convolution greatly broaden the feature's receiving held to catch any connections between long-distance characteristics, whilst the four layers of normal convolution can identify the image's local features. It just takes a 7-layer convolution to get a 26 x 26 × 26 reception.

Table 1's 3D U-Net-SparseVoxNet, which DenseNet inspired, shows a skip link as a black dotted line. Using deconvolution methods to bypass the connection, the image gets split once. The network will converge faster thanks to the skip connection and to have a greater accuracy rate. Because the narrow neural network gains more edge information and loses less information via convolution, the first segmented picture is going to perform better on edge segmentation. A fine-grained segmentation is the end outcome. In terms of coarse-grained segmentation, the second segmented picture performs better overall.

Table 1. Each layer's stride and convolution kernel

Input Image	Output	Layer	Stride	Kernel	Parameters
64 64	32 32	Conv_1	2	3	448
32 32	16 16	Conv_2	2	3	6928
16 16	16 16	Spatial attention	2	1	816
16 16	16 16	Sparse Block_1	1	3	43330
16 16	16 16	Sparse Block_2	1	3	496984
128	128	Conv_3	2	3	11840
2	2	Conv_4	1	3	6464

Fig. 1 Proposed architecture model

High-level abstract characteristics from deep neural networks are very useful for removing the whole tissue's divided central portion. By voting, the last segment's outcome is decided by several segmentation outcomes of various clipped data shown in a single voxel. Sparse blocks take on the role of U-Net's down sampling method, and both deconvolutions are comparable to the up-sampling procedure.

The number of parameters is shown in Table 1 for each layer in the SparseVoxNet. The parameters of two deconvolution layers, two sparse blocks, a skip connection layer, a spatial attention mechanism layer, and four convolution layers are shown in Table 1. Conv_n represents the four convolution layers, Deconv_n represents the two deconvolution layers, and Sparse Block_n represents the two sparse blocks.

Figure 1 depicts the SparseVoxNet architecture that is suggested in this work. It improves U-Net, which performs end-to-end training via the use of up sampling and down sampling processes. Dilated convolutions are used in the last three layers, and regular convolutions in the first four levels. There are three, four, and five holes. To reinforce the most significant aspects of the first feature map, following the data feature map is the implementation of the spatial self-attention mechanism. A very precise segmentation is the product. In terms of coarse-grained segmentation, the second segmented picture performs better overall. High-level abstract characteristics from deep neural networks are very useful for extracting the segmented centre region of the whole tissue.

We also count the number of parameters in each layer of the SparseVoxNet that Table 1 displays. Table 1 displays the settings for four convolution layers: a skip connection layer, two sparse blocks, a spatial attention mechanism layer, and two deconvolution layers. Conv_n represents the four convolution layers, Deconv_n represents the two deconvolution layers, and Sparse Block_n represents the two sparse blocks. Table 1 also displays the stride and convolution kernel for each layer. Keep in mind that every entry in Table 1 matches every layer in Figure 1.

3.2. Sparse Block

Compared to ResNet, DenseNet has denser connections, which results in much higher hardware resource usage. We, therefore propose sparse network architecture to alter feature reuse while preserving feature reuse in addition to skipping connection features. By only offering direct connections, often referred to as full skip connections, between a pair of layers that share the same feature map size, our suggested sparse blocks minimize the overall number of connections. Nonetheless, the sparse block's influence is equivalent to the dense blocks. The following is the transition layer's input:

$$\left[T_0, T_1, T_2, T_3, T_4\right] = \left[T_0, H_1(T_0), H_2(T_1), H_3(T_2), H_4(T_3)\right] \quad (1)$$

Where the input of H_1 is T_0, the input of H_2 is $T_0 + T_1$ and soon. Different scales relate to the feature mappings of various receptive fields. It has been discovered that the linear combination outperforms the nonlinear combination when it comes to combining the characteristics of various scales. Composite expressive Feature maps of different sizes are immediately layered to build features; this method was inspired by the architecture of the U-Net network.

In contrast to U-Net, the enhanced network architecture replaces the up-sampling step with deconvolution, minimizing information loss in the conversion process. Overfitting in Dense Net may be readily caused by the excessive density of the network connections between the preceding layer and the later layer. The sparse network can resolve this issue. There is no disappearing gradient, and the network can represent features much better.

Wang et al. [22] suggested a nonlocal block, a self-attention technique, for capturing long-range dependence, which was inspired by non-local mean filtering for images. Nonlocal blocks compute the connection between two places directly, disregarding the Euclidean distance. It computes the features' generalized autocorrelation matrix. Nonetheless, there is a fair amount of computation efficiency. Since the network's fitting ability may be achieved without stacking too many deep convolution processes following the addition of nonlocal operators, the first sparse block is preceded by the spatial self-awareness model since it can be readily incorporated into the network and does not alter the amount of input data. In this study, we implement the self-attention method as suggested by Zhang et al. [23]. The 3D network, which has the following definition, embeds the nonlocal block:

$$yi = \frac{1}{C(x)} \sum_{\forall j} f\left(x_i, x_j\right) g\left(x_j\right) \quad (2)$$

Where x is the input data, j is an index of all possible locations, i is a 3D coordinates that shows the current location index of the incoming data, q is a unary conversion function, and f is an asynchronous computation function that may ascertain the relationship between the i^{th} and j^{th} locations. In the experiments, the multichannel feature is fused, and the ascending dimension is achieved using the $1 \times 1 \times 1$ convolution; C (x) is then utilized for normalization. In the attention model, the number of channels may be decreased or increased while still achieving cross-channel interactions and information integration via the use of multiple one × one × one convolution kernel. Due to Lin's proposed network structure [24], the $1 \times 1 \times 1$ convolution came to people's notice, which links two complete connection layers for linear feature fusion. Next, the inception module of the $1 \times 1 \times 1$ Google's Inception-v4 [25] network design uses convolution for dimensional reduction or ascending dimension. Motivated by such benefit, the $1 \times 1 \times 1$ convolution kernel is utilized in this work to determine the spatial autocorrelation connection, lower the dimensionality of the initial input data, and subsequently increase the data's dimension. The distinct computed weights are appended to the initial data and afterwards regularised to depict the impact on attributes of voxels situated in disparate spatial orientations.

4. Experiments and Results

4.1. Dataset

4.1.1. ACDC

The Automated Cardiac Diagnosis Challenge (ACDC) dataset, which contains CMR pictures from 150 patients, gives a much bigger and more adjusted dispersion of distinctive heart conditions, counting ordinary, infracted, and cardiomyopathic hearts. It incorporates comments for both the myocardium and ventricles, making it a flexible asset for a wide run of cardiac division assignments. Be that as it may, ACDC needs the center on CHD found in HVSMR 2016. For CHD-specific inquiries about, HVSMR 2016 offers more specialized information, indeed in spite of the fact that ACDC's bigger measure and differences make it stronger for common division challenges.

4.1.2. UK Biobank

UK Biobank offers cine MRI information from over 100,000 members, giving one of the biggest cardiovascular datasets accessible. Its endless estimate makes it perfect for profound learning applications, especially when creating models that can be generalized over diverse socioeconomics and conditions. Be that as it may, UK Biobank needs the nitty gritty blood vessel and myocardium comments displayed in HVSMR 2016, and it does not particularly center on CHD, making it less valuable for analysts in this specialized range.

4.1.3. Sunnybrook Cardiac Information (SCD)

The Sunnybrook dataset, which comprises cine MRI pictures from 45 patients, centers essentially on clearing out the ventricular division. Its point by point explanations and consideration of both solid and obsessive cases make it perfect for cleared out ventricular investigation, but it needs the broader scope of explanations accessible in HVSMR 2016. For complex anatomical considerations that require point by point division of the myocardium and blood vessels, HVSMR 2016 is more suitable, indeed, in spite of the fact that Sunnybrook may offer superior execution in cleared out ventricular-specific tasks.

4.1.4. MyoPS

MyoPS gives multi-sequence CMR information that provides a wealthier see of tissue characteristics, making it perfect for myocardial pathology discovery. In any case, MyoPS does not incorporate a focus on CHD or blood vessel division, regions where HVSMR 2016 exceeds expectations. For analysts fascinated by CHD-specific ponders, HVSMR 2016 remains the more important asset, whereas MyoPS is superior suited for considerations centered on tissue characteristics and pathology discovery.
MICCAI 2009 Cleared out Ventricle Division Challenge: Like Sunnybrook, the MICCAI 2009 dataset centers on cleared out ventricle division, advertising point by point comments for the endocardium and epicardium. Whereas HVSMR 2016 offers broader division conceivable outcomes, counting the myocardium and blood vessels, MICCAI 2009 is more focused on assets for cleared out ventricular division assignments. Both datasets are moderately small, making information increase pivotal in preparing machine learning models.

There are six primary data formats for radiobiological imaging. Among them is the Neuroimaging Informatics Technology Initiative, or NIFTI. To link a voxel's physical index to its real spatial position, this format includes two affine coordinates. We evaluate the network design and the approach using the HVSMR 2016 dataset. A total of ten training images and ten testing scans for cardiac magnetic resonance are available with HVSMR 2016. The myocardium and main blood vessels are annotated in each cardiac 2D echo training set, which is derived from patients with a diagnosis of Complicated Heart Disease (CHD).

All the cardiac MR images have been normalized due to the significant intensity differences among the images. The mean and unit variance are zero after normalization. Simple augmentation of data was used to increase the training data to make use of the restricted training set. Cropping and rotation are examples of augmentation activities. The three parts that comprise the first training set are the validation set, testing set, and training set. In parameter training, the cross-validation procedure is used. Of the photos, 70% are used for training, and 30% are used for testing. After that, we contrast and quickly go through the experimental findings.

4.2. Evaluation Metrics

A crucial stage in the processing of medical images is segmentation. On the other hand, choosing an appropriate assessment index to compare segmented medical images and assess segmentation quality is challenging. This research employs the following three measures to assess the segmentation outcomes:

4.2.1. Dice Coefficient

A common method for assessing the effectiveness of 3D medical picture segmentation is the dice coefficient. Ensuring good recall and accuracy is the main goal. The segmentation impact may be more accurately quantified by employing the Dice coefficient assessment approach as opposed to the direct computation of the difference among the automated segmentation results as well as the original data labels. The definition of a dice coefficient is:

$$Dice = \frac{2|G \cap R|}{|R| + |G|} = \frac{2TP}{2TP + FP + FN} \quad (3)$$

Where G is a segmentation outcome using the labelled testing data or ground truth, R is the test data's automated segmentation result. For every class, TP, FP, and FN stand for True Positives, False Positives, and False Negatives, respectively. The segmentation result template and the label data template should ideally fully overlap, meaning that R = G and the Dice coefficient's absolute value equals 1.

4.3. Training

The investigations use the unsupervised gradient-descending optimization approach and randomly initialize all weights using the Gaussian distribution. The batch consists of eight pieces. The load absorption is adjusted to 0.0005, and acceleration is increased to 0.9 to speed up training. This combination facilitates escaping from extreme locations and prevents one from being trapped in locally optimum solutions. These parameters help the model become less overfit and accelerate its convergence.

The drop rate is set at 0.2, and the starting learning rate to 0.01. The drop rate is degraded and reinitialized every 5000 steps, with a somewhat large initial drop rate. If the rate of learning is too high, the model is going to be unstable as well as never converge. A Dual RTX 2080 Ti GPU was used for training and testing our algorithms.

The impact of the enhanced approach on segmentation is intended to be verified by many sets of ablation tests and comparison studies. In the trials, we contrasted our approach with other deep learning techniques and conventional techniques. We also examined the network with the mechanism for attention as well as DenseVoxNet alone, as well as the network with mixed dilated convolution.

Model definition: 3D U-Net-SparseVoxNet-S. When the spatial system that drives self-attention is included, this is the model. The findings demonstrate that the average symmetrical surface distance (ADB) and the myocardium's Hausdorff distance outperform DenseVoxNet. The heart and blood pool has Hausdorff distances that are around 3.0% and 4.8% greater than DenseVoxNet. This implies that the spatially self-attention mechanism has been effective in bringing the segmented pictures closer to their target domain. Our model uses self-attention in addition to convolution to represent global-level, long-range relationships inherent in cardiac structure. The attention mechanism has the following benefits: (a) minimal parameters, (b) quick computation, and (c) the ability to capture features at a distance. Because of the short sample training approach used in this work, the segmentation outcome is not optimal when a spatial self-attention process is eliminated, making it difficult to extract long-range features effectively. We integrate both dilated convolution as well as the stretched convolution-based technique since it only collects information from a limited number of surrounding locations and is unable to give rich context information, spatial self-attention approaches are used to acquire long-range characteristics. A single feature at any place could be able to see the characteristics of every other site due to the spatial focus process itself, which might result in more potent pixel-level representation capabilities. These results suggest that spatial self-attention combined with 3D dilated convolution greatly facilitates the usage of multistage data.

4.4. Results

Figure 2 displays the segmentation results for three training photos. These three pieces come from various people. The data in the example dataset with an index of 60 have an analogous single-dimensional coronal plane viewpoint. The first line's image depicts the myocardium and blood pool in both bright and dark blue regions, with the background consisting of black and dark blue elements. The described photographs in the next line correspond to the myocardium as well as the plasma pool in the initial row of pictures. The automated segmentation results utilizing the work's technique are shown in the third line. The colours dark purple, yellow, and blue represent the blood pool, the backdrop, and the myocardium, respectively. Using a low-contrast cardiac 2D echo, Figure 3 shows that even with significant differences in the cardiac architecture of the training group members, our proposed method may effectively alter the myocardial and blood pool. This proves that the method can accurately match the initial data. There continue to be some issues, however. The bottom left corner of the first auto-segmentation result shows a partial

myocardial separation. The myocardium showed the backdrop in the second finding. The extra heart muscle on the top right corner of the third result demonstrates that deep learning can recognize many of the characteristics in the data but is not capable of making good logical deductions. These minute logical mistakes cannot be produced by human segmentation.

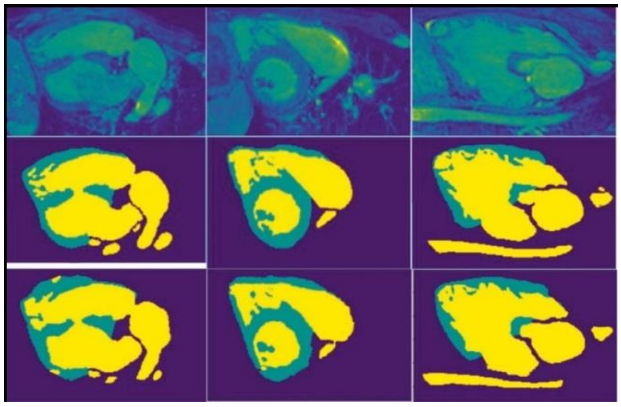

Fig. 2 Outcomes of segmentation using three training photos

Three test photos' segmentation results are shown in Figure 1. The technique of data extraction remains unchanged. The strategy presented in this research also has a strong generalization impact on unlabelled data, as we can see by looking at the outcomes. However, due to the large number of parameters, the gradient descent technique may easily lead to overfitting by entering the local optimum.

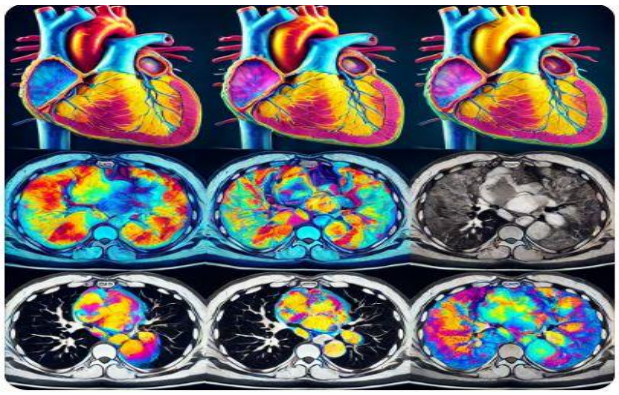

Fig. 3 Results of segmentation on three test images

4.5. Discussion

Figure 3 displays a comparison of the outcomes from the six previous ways and the one we suggested. Their primary ranking is based on the Dice coefficient. The image displays two instances of supplementary reference indices: the symmetric Hausdorff distance and ADB. The last three deep learning techniques utilize the HVSMR 2016 Challenge dataset, whereas the initial three use more conventional techniques like feature extraction by hand and hidden Markov randomized fields. As a result of the myocardium's hazy borders during low-resolution 2D echo, Figure 3 demonstrates that segmenting the blood pool is much easier since the blood pool's Dice coefficient is larger than the myocardial in all methods. In terms of myocardial segmentation, our suggested approach performs the best when using the Dice; that is, the challenge's ranking measure, 0.861 ± 0.024, beats the second one by almost 4%. The highest performance was likewise attained in blood pool segmentation using Dice; the challenge's ranking score of 0.94 ± 0.016 shows that our sparse interconnected network can take on the challenging cardiovascular segmentation task. Our method's Harsdorf distance and ADB also produced the best results.

Table 3 primarily displays the results of additional 3D and 2D echo segmentation techniques. First, a comparison of the experimental parameters reveals that the approach suggested in this work requires the fewest parameters. With the inclusion of the attention model, the sparse block, as well as dilated convolution, may produce a number of factors. In many situations, the dense block's feature expression ability will outperform the sparse blocks. However, in this paper, the sparse block is applied to the problem of medical segmentation and small sample training, which allows it to fit and generalize the data well with fewer parameters. Additionally, the convolution operations are reduced by the dilated convolution's exponentially increasing receptive field. The attention mechanism is capable of effectively capturing the elements that enhance the network's capacity for generalization. The model has a quick convergence rate and requires less computation due to its minimal number of parameters.

Table 2. Experimental findings are compared between the 3D approaches and the enhanced method

Method	Dice	ABD	Hausdorff
3D U-Net	74.26	2.412	12.36
VoxResNet[19]	79.62	2.341	7.25
3D U-Net-3D U-Net-SparseVoxNet	89.92	0.751	4.36

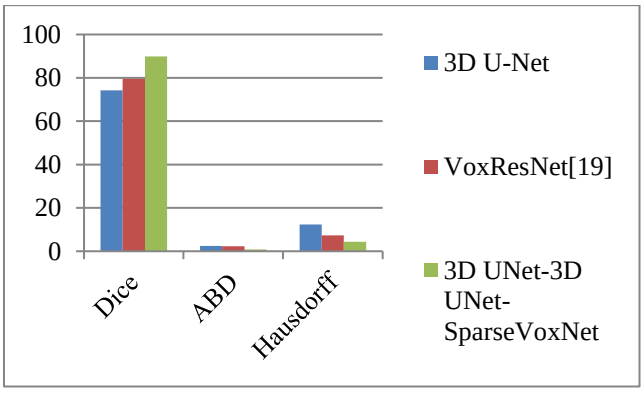

Fig. 4 Comparison of the upgraded method's experimental findings with those of other approaches

5. Conclusion

To separate the myocardial and blood pool from 2D echo images, we suggest an innovative and effective 3D sparse convolution network in this research. This technique may lower model parameters, get rid of pointless computations, and lessen the chance of overfitting data used for training on small samples. The ability of feature maps to

be expressed may be optimized by the spatial self-attention mechanism, and the convolution network depth can be decreased by sparse blocks. This study presents a precise pixel-by-pixel categorization. Additionally, we get competitive outcomes when compared to current methodologies. The suggested approach may provide medical professionals with all the information they need to diagnose CHD.

References

[1] Michaela Schmidt et al., "Novel Highly Accelerated Real-Time CINE-MRI Featuring Compressed Sensing with K-T Regularization in Comparison to TSENSE Segmented and Real-Time Cine Imaging," *Journal of Cardiovascular Magnetic Resonance*, vol. 15, pp. 1-2, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[2] Michael S. Hansen et al., "Retrospective Reconstruction of High Temporal Resolution Cine Images from Real-Time MRI Using Iterative Motion Correction," *Magnetic Resonance in Medicine*, vol. 68, no. 3, pp. 741-750, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[3] Li Feng et al., "5D Whole-Heart Sparse MRI," *Magnetic Resonance in Medicine*, vol. 79, no. 2, pp. 826-838, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[4] Simone Coppo et al., "Free-Running 4D Whole-Heart Self-Navigated Golden Angle MRI: Initial Results," *Magnetic Resonance in Medicine*, vol. 74, no. 5, pp. 1306-1316, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[5] M. Usman et al., "Free Breathing Whole-Heart 3D CINE MRI with Self-Gated Cartesian Trajectory," *Magnetic Resonance in Medicine*, vol. 38, pp. 129-137, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[6] Cameron Hassani et al., "Myocardial Radiomics in Cardiac CMR," *American Journal of Roentgenology*, vol. 214, no. 3, pp. 536-545, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[7] Yucheng Song et al., "Deep Learning-Based Automatic Segmentation of Images in Cardiac Radiography: A Promising Challenge," *Computer Methods and Programs in Biomedicine*, vol. 220, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[8] Daniel Rueckert, Gabor Fichtinger, and S. Kevin Zhou, *Handbook of Medical Image Computing and Computer Assisted Intervention*, Elsevier, 2020. [Google Scholar] [Publisher Link]

[9] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning*, MIT Press, pp. 1-800, 2016. [Google Scholar] [Publisher Link]

[10] Bosung Seo et al., "Cardiac MRI Image Segmentation for Left Ventricle and Right Ventricle Using Deep Learning," *arXiv*, pp. 1-27, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[11] David Chen et al., "Deep Neural Network for Cardiac Magnetic Resonance Image Segmentation," *Journal of Imaging*, vol. 8, no. 5, pp. 1-11, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[12] Kamalika Some, The History, Evolution and Growth of Deep Learning, Analytics Insight, 2018. [Online]. Available: https://www.analyticsinsight.net/the-history-evolution-and-growth-of-deep-learning/

[13] Xiaofeng Ding et al., "Cab U-Net: An End-to-End Category Attention Boosting Algorithm For Segmentation," *Computerized Medical Imaging and Graphics*, vol. 84, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[14] S. Jeevakala et al., "Artificial Intelligence in Detection and Segmentation of Internal Auditory Canal and Its Nerves Using Deep Learning Techniques," *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, no. 11, pp. 1859-1867, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[15] Geert Litjens et al., "A Survey on Deep Learning in Medical Image Analysis," *Medical Image Analysis*, vol. 42, pp. 60-88, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[16] Fisher Yu, and Vladlen Koltun, "Multi-Scale Context Aggregation by Dilated Convolutions," *arXiv*, pp. 1-13, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[17] Jelmer M. Wolterink et al., "Dilated Convolutional Neural Networks for Cardiovascular MR Segmentation in Congenital Heart Disease," *Reconstruction, Segmentation, and Analysis of Medical Images: First International Workshops*, Athens, Greece, pp. 95-102, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[18] Fisher Yu, Vladlen Koltun, and Thomas Funkhouser, "Dilated Residual Network," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 472-480, 2017. [Google Scholar] [Publisher Link]

[19] Hao Chen et al., "VoxResNet: Deep Voxelwise Residual Networks for Brain Segmentation from 3D MR Images," *NeuroImage*, vol. 170, pp. 446-455, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[20] Lequan Yu et al., "Volumetric ConvNets with Mixed Residual Connections for Automated Prostate Segmentation from 3D MR Images," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, pp. 66-72, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[21] Lee R. Dice, "Measures of the Amount of Ecologic Association between Species," *Ecology*, vol. 26, no. 3, pp. 297-302, 1945. [CrossRef] [Google Scholar] [Publisher Link]

[22] Kolawole O. Babalola et al., "Comparison and Evaluation of Segmentation Techniques for Subcortical Structures in Brain MRI," *Medical Image Computing and Computer-Assisted Intervention: 11th International Conference*, New York, NY, USA, pp. 409-416, 2008. [CrossRef] [Google Scholar] [Publisher Link]

[23] Vasily Zyuzin et al., "Identification of the Left Ventricle Endocardial Border on Two-Dimensional Ultrasound Images Using the Convolutional Neural Network Unet," *2018 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology*, Yekaterinburg, Russia, pp. 76-78, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[24] Ozan Oktay et al., "Anatomically Constrained Neural Networks (Acnns): Application to Cardiac Image Enhancement and Segmentation," *IEEE Transactions on Medical Imaging*, vol. 37, no. 2, pp. 384-395, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[25] Özgün Çiçek et al., "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," *Medical Image Computing and Computer-Assisted Intervention: 19th International Conference*, Athens, Greece, pp. 424–432, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[26] Olivier Bernard et al., "Standardized Evaluation System for Left Ventricular Segmentation Algorithms in 3D Echocardiography," *IEEE Transactions on Medical Imaging*, vol. 35, no. 4, pp. 967-777, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[27] Jianxu Chen et al., "Combining Fully Convolutional And Recurrent Neural Networks For 3d Biomedical Image Segmentation," *Advances in Neural Information Processing Systems*, vol. 29, 2016. [Google Scholar] [Publisher Link]

[28] M. Ghafoorian et al., "Non-Uniform Patch Sampling with Deep Convolutional Neural Networks for White Matter Hyperintensity Segmentation," *IEEE 13th International Symposium on Biomedical Imaging*, Prague, Czech Republic, pp. 1414-1417, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[29] Rudra P.K. Poudel, Pablo Lamata, and Giovanni Montana, "Recurrent Fully Convolutional Neural Networks for Multi-Slice MRI Cardiac Segmentation," *Reconstruction, Segmentation, and Analysis of Medical Images: First International Workshops*, Athens, Greece, pp. 83-94, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[30] Jonathan Long, Evan Shelhamer, and Trevor Darrell, "Fully Convolutional Networks for Semantic Segmentation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431-3440, 2015. [Google Scholar] [Publisher Link]

[31] Tsung-Yi Lin et al., "Feature Pyramid Networks for Object Detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2117–25, 2017. [Google Scholar] [Publisher Link]

[32] Sarah Leclerc et al., "Deep Learning for Segmentation using an Open Large-Scale Dataset in 2D Echocardiography," *IEEE Transactions on Medical Imaging*, vol. 38, no. 9, pp. 2198-2210, 2019. [CrossRef] [Google Scholar] [Publisher Link]