MMC-Net: Multi-modal network for cardiac MRI segmentation of ventricular structures, and myocardium

G. Jignesh Chowdary, Pratheepan Yogarajah, *Member, IEEE* and Priyanka Chaurasia

Abstract—Automatic segmentation of multi-modal Cardiac Magnetic Resonance Imaging (CMRI) scans is challenging due to the variant intensity distribution and unclear boundaries between the neighbouring tissues and other organs. The deep convolutional neural networks have shown great potential in medical image segmentation tasks. In this paper, we present a deep convolutional neural network model named Multi-Modal Cardiac Network (MMC-Net) for segmenting three cardiac structures namely right ventricle (RV), left ventricle (LV), and left ventricular myocardium (LVM) from multi-modal CMRI's. The proposed MMC-Net is designed using a densely connected backbone enabling feature reuse, an atrous convolution module for fusing multi-scale features, and a pixel-classification module for generating the segmentation result. This model was evaluated on a publicly available MS-CMRSeg-2019 challenge dataset in segmentation of RV, LV, and LVM from CMRI scans. The segmentation results from extensive experiments demonstrate our MMC-Net can achieve better segmentation performance compared to other state-of-the-art models, and the existing approaches. Additionally, the generalization ability of the proposed MMC-Net is validated on another publicly available ACDC dataset without fine-tuning. The results demonstrate that the proposed MMC-Net shows a powerful generalisation ability of segmenting RV, LV, and LVM with higher performance.

Index Terms—CMRI, atrous convolution, MS-CMRSeg-2019, cardiac structures, multi-modal, ACDC.

1 Introduction

Cardiac Magnetic Resonance Imaging (CMRI) is a non-invasive imaging procedure used to diagnose cardiovascular diseases. In CMRI, the Left ventricle (LV), Right Ventricle (RV), and Left ventricular Myocardium (LVM) are obtained for evaluating the type of cardiac disease and potential threats [1]. CMRI is used to monitor cardiac health, especially in cases where patients are suffering from arrhythmia and facing issues of holding thriller breath. Different pulse sequences like Late Gadolinium enhancement (LGE), balanced steady-state free precession (bSSFP), T2-Spectral Presaturation Attenuated Inversion-Recovery (T2-SPAIR) can help to distinguish normal anatomy from various pathologies. The LGE CMR is capable of highlighting the infarcted myocardium with distinctive brightness over the healthy myocardium [2], [3]. The T2-SPAIR CMR is a black-blood technique sensitive to myocardium edema. The bSSFP CMR can obtain clear ventricular boundaries and capture cardiac motions. But the manual delineation of the cardiac structures (LV, RV, LVM) is tedious, time-consuming, subject to inter and intra variations [2], [4], and often prone to human error. Even for experienced medical professionals, it is very challenging to delineate the LVM, LV, and RV from the CMR images [5], [6]. Thus, automation of this procedure is the need of the hour in clinical practice. Fig. 1 shows sample of of LGE, bSSFP and T2-SPAIR CMRI slices and ground truth region of LV, RV, and LVM.

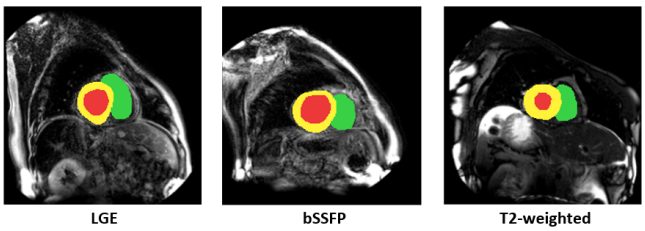

Fig. 1. Sample of LGE, bSSFP and T2-SPAIR CMRI slices; the red, green, and yellow regions represents the ground truth of LV, RV, and LVM.

However the automatic segmentation of LVM, LV, and RV from the CMR images is challenging due to motion artifacts, ambiguous boundaries between the neighboring tissues and organs, non-uniform intensity distribution, and shape variations [7]. The LGE, and bSSFP CMR images have same intensity distribution, whereas for T2-SPAIR CMRI, the intensity distribution is different. The existing automatic image segmentation approaches can be divided into two categories, namely conventional image processing (atlas-based and prior knowledge-based methods) and deep learning methods. The prior knowledge-based methods [8], [9], [10], [11], [12], [13], [14] acquire the prior shape information from the corresponding CMR images of the same patient for segmentation; these methods require inter-modality registration of CMR images. Atlas-based methods rely on 3D

- *GJ. Chowdary was an under graduate student at Vellore Institute of Technology, Chennai, India. E-mail: jigneshchowdary@gmail.com*
- *Pratheepan Yogarajah was a Lecturer in computer science at Ulster University, Northern Ireland, UK. Email: p.yogarajah@ulster.ac.uk*
- *Priyanka Chaurasia was a Lecturer in data science at Ulster University, Northern Ireland, UK. Email: p.chaurasia@ulster.ac.uk*

Manuscript received xxxxx xxx, 2022; revised xxxxxx xx, 2022.

or 2D image registration to align the images from different image modalities or patients. These inter-subject and inter-modality image registrations are very challenging problems.

The deep learning-based methods employ deep convolutional neural networks (DCNN) to segment regions of interest. The Fully Connected Network (FCN) [15] is a representative DCNN that has achieved good results in semantic segmentation and bio-medical segmentation tasks. This network has been adopted by [16] for segmenting the cardiac regions in CMR images. Another representative DCNN is U-Net [17]; this network is widely used as a baseline in many biomedical segmentation tasks, including cardiac structure segmentation. Li et al. [18] designed a dilated inception module and embedded it in U-Net to extract multi-scale features from CMR images for RV segmentation. Painchaud et al. [19] designed a network that includes a variational auto-encoder (VAE) and an anatomical post-processing module based on U-Net for cardiac image segmentation. Okasuz et al. [20] designed a time-DCNN network for the detection of motion artifacts, reconstruction, and segmentation of cardiac structures from CMRI.

These methods achieved satisfactory performance when CMRI slices were not closely related. A 3D-DCNN is a good solution for slices with tight correlation. The Recurrent Neural Networks (RNNs) are another option for processing successive MRIs. Poudel et al. [21] designed a recurrent fully-convolutional network for multi-slice CMRI segmentation. For capturing the important dynamics of cardiac motion, Zhang et al. [22] developed a multi-stage convolutional LSTM for segmenting LVM from infarcted porcine cine MR images. But these 3D networks are computationally expensive, so researchers frequently utilize 2D CNN for CMR segmentation. The above-discussed methods are prone to over-segmentation or under-segmentation of cardiac structures.

We present an atrous convolution-based model named Multi-Modal Cardiac Network (MMC-Net) for CMRI segmentation to improve the segmentation performance. We evaluated the proposed segmentation model on the MICCAI CMR segmentation challenge (MS-CMRSeg-2019) dataset and the ACDC dataset. The MS-CMRSeg consists of CMR sequences (LGE, T2-SPAIR, and bSSFP) from 45 patients. The ACDC dataset consists of cine CMR sequences collected from 100 patients. The main contributions of this work are:

- The proposed MMC-Net can segment the RV, LV, and myocardium volume or LVM from short-axis CMR images with a high degree of precision and recall.
- The proposed MMC-Net does not require image registration, complicated pre/post-processing, or a complex training strategy. It is trained with the multi-modality directly.
- The proposed MMC-Net demonstrates a powerful generalization capability without any fine-tuning.
- On the MS-CMRSeg and ACDC dataset, the proposed MMC-Net achieved higher performance than the state-of-the-art methods and existing works.

The rest of the paper is organized as follows, Section 2 describes the methodology, Section 3 presents the results and discussion of the experiments, and Section 4 concludes the proposed work.

2 METHODOLOGY

The proposed MMC-Net is shown in Fig. 2. This model consists of a densely connected backbone, an atrous convolution module (ACM), and a pixel classification module. The densely connected backbone consists of nine convolutional layers, where the first convolutional layer is used to extract initial features, and the rest eight convolutional layers are connected densely. The ACM is integrated with the backbone for learning multi-scale contextual features of the cardiac structures. Finally, the extracted features are passed to the pixel-classification module, where a set of two 1×1 convolution layers followed by a soft-max layer are used for generating the segmentation result. The densely connected backbone, and the ACM module are explained below.

2.1 Densely connected backbone

In this work, we employed dense connectivity [23] to design our backbone network to promote the information flow among the layers and effectively reuse the features. In the backbone, each layer is connected to the succeeding layers. The group of convolution layers with this connectivity pattern is called dense block, as shown in Fig. 2. The output of the dense block is defined as follows:

$$Output_{DB} = \sigma(y_o) + \sigma(y_1) + ... + \sigma(y_n) \tag{1}$$

In Equation 1, the σ represents the non-linear mapping of convolutional layers followed by batch-normalization and ReLU layers, $y_0, y_1, ..., y_n$ represents the input to the layers in the dense block, and $+$ represents the concatenation. In this work, as we have used only one dense block, i.e., $DB = 1$. The input of the last layer (y_n) in the dense block is a concatenated result of all the preceding layer's output of non-linear mapping, as shown below:

$$y_n = \sigma(y_0) + \sigma(y_1) + ... + \sigma(y_{n-1}) \tag{2}$$

This indicates that the layers in the dense block are fully connected, and the feature maps have a consistent size. Each non-linear mapping outputs m feature maps, where m is a hyper-parameter known as growth rate. In other words, m can be expressed as the number of filters for each convolutional layer in the dense block. For instance, let's consider the m_0 as the input feature map to the dense block, then the last layer (n^{th} layer) of the dense block has $m_0 + m \times (n-1)$ feature-maps as input.

To summarise, the backbone employed in this work uses dense connectivity for learning discriminative features. The proposed backbone uses skip connections to deal with the problem of vanishing gradients as the network goes deeper. In addition, this dense connectivity employs concatenation operations to aggregate the network's features. This ensures the flow of information between the layers, which permits the model to reuse the feature maps outputted by the preceding layers to retain the highly-discriminative local features.

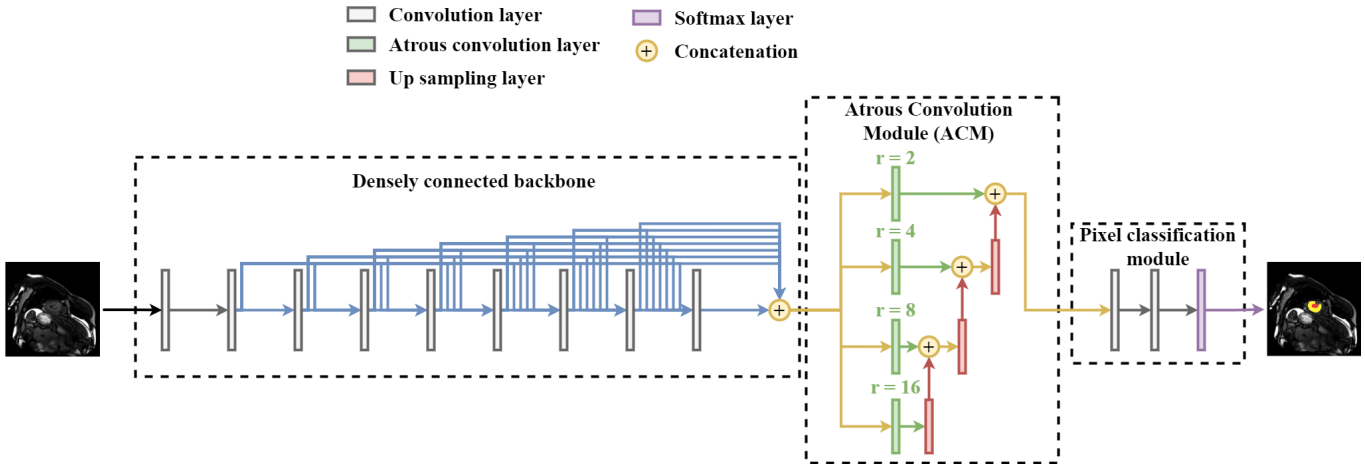

Fig. 2. Architecture of the proposed MMC-Net.

2.2 Atrous Convolution Module

Atrous convolutions resolve the issue of reduced resolution (by the CNN) by adjusting the filter's receptive field. The atrous convolution works by inserting holes (i.e., zeros) between the pixels in the convolutional kernels. This was initially used to compute the wavelet decomposition [24], and undecimated wavelet transform [25] and is also known as the "algorithm atrous" [26]. Consider an example of a one-dimensional signal with input signal as $x[i]$ and output signal as $y[i]$ for every location of i, the atrous convolution can be defined as:

$$y[i] = \sum_{k=1}^{K} x[i + r \cdot k]a[k] \tag{3}$$

In Equation 3, the r is the dilation fraction and $a[k]$ is the kernel with the length of k. When $r = 1$, the atrous convolution is the same as the standard convolution. When $r = 1$, $stride = 1$, and $padding = 0$, the operation is similar to downsampling. The CNN's can learn dense features when the stride is smaller, and padding approximates the dilation rate. The density of the feature maps extracted by the CNN's can be controlled by adjusting the dilation rate and stride. Assuming a 2D standard convolution of size $c \times c$, the corresponding atrous convolution size c_{atrous} can be calculated using the formula below:

$$c_{atrous} = c + (c - 1) \times (r - 1) \tag{4}$$

Equation 4 indicates that the input signal is not changed, but it is just upsampled by inserting $r - 1$ zeros between the pixels. The size of the receptive field of $(z - 1)^{th}$ layer be denoted as R_{z-1}, the receptive field of the z^{th} 2D atrous convolution can be computed as:

$$R_z = R_{z-1} + [c + (c - 1) \times (r - 1))] \times T_z \tag{5}$$

In Equation 5, r is the dilation fraction, T is the stride of the z^{th} layer. The stride of the z^{th} layer can be computed as the product of the preceding strides. For a $2D$ image I, the output feature map Y at the z^{th} layer can be computed as follows:

$$Y = (I + 2P_d - R_{z-1})/T_z + 1 \tag{6}$$

Where P_d is the padding of the atrous convolution.

In this work, we design a novel module known as Atrous Convolution Module (ACM) that is added at the end of the backbone to promote multi-scale feature learning. This module employs multi-input-parallel-branch structure as shown in Fig. 2, with four atrous convolutions in parallel with different dilation rates. In ACM we also employed up-sampling layers between atrous convolutions and concatenation at each branch to keep the dimensions of the feature-maps consistent with the previous branch.

2.3 Loss function

In this work, we employed a weighted dice loss function to deal with the complex nature of cardiac structure segmentation on CMR images. This loss function is illustrated as follows:

$$Dice_{weighted} = 1 - D_{weighted} \tag{7}$$

In Equation 7, $D_{weighted}$ is the weighted dice score and it can be computed as follows:

$$D_{weighted} = \frac{2\sum_p w_p D_p}{\sum_p w_p} \tag{8}$$

In Equation 8, the w_p indicates the weight assigned to class p. We have empirically assigned 4, 4, 4, and 1 for RV, LV, LVM, and background in this work. We assigned a lower weight to the background because it covers a large portion of the CMRI slice than the other three structures (i.e., RV, LV, LVM), and the other three structures are assigned equal weights because they were given equal importance. Higher weights are assigned to force the segmentation model to emphasize more on these structures during training. D_p is the dice score for class p, and it can be computed as follows.

$$D_p = \frac{2\sum g_i \hat{y}_i}{\sum g_i + \sum \hat{y}_i} \tag{9}$$

In Equation 9, $\hat{y}_i$ is the predicted output of the proposed MMC-Net and g_i is the ground-truth.

3 Experiments and Results

3.1 Dataset

The proposed MMC-Net is trained and evaluated using the MS-CMRSeg 2019 challenge [2], [3]. For understanding the generalization ability of the MMC-Net we additionally used a publicly available dataset from different imaging modality namely the Automated Cardiac Diagnosis Challenge (ACDC) [27].

The MS-CMRSeg-2019 dataset consists of MRI data from 45 cardiomyopathy patients, each scanned using LGE, T2-SPAIR, and bSSFP sequences from short-axis orientation. The LGE sequences have more than ten slices covering the entire ventricular region with a resolution of 5 mm in slice and $0.75 \times 0.75\ mm$ in-plane thickness. The T2-SPAIR sequences are very short, with only 4 to 7 slices, and their image resolution is about 12 to 20 mm in slice and $1.35 \times 1.35\ mm$ in in-plane thickness. The bSSFP sequences cover the ventricular region from the apex to the basal plane of the mitral valve with a resolution of 8 to 13 mm in slice and $1.25 \times 1.25\ mm$ in-plane thickness. During the challenge, the manual annotations are provided only for a subset of scans, which includes 35 scans of T2-SPAIR and bSSFP sequences and only 5 scans of LGE sequences. The remaining scans are used for evaluation and are not released. But recently the challenge organisers provided the gold-standard annotations for the remaining scans, which are used for evaluation in this work.

The ACDC dataset is created by the University Hospital of Dijon (France). This dataset consists of short-axis cine CMRI scans acquired from 150 subjects in both end diastole (ED), and end systole (ES) phases. These scans are taken using two MRI scanners with different weighting strategies (1.5T and 3T) respectively. These scans are obtained from healthy patients and also from patients with abnormal right ventricle (ARV), hypertrophic cardiomyopathy (HCM), dilated cardiomyopathy (DCM), and previous myocardial infarction (MINF). These cine CMRIs have a slice thickness of 5 to 9 mm, inter-slice gap of 5 or 10 mm, and spatial resolution of 1.37 - 1.68 mm^2. In this dataset, the segmentation ground-truths are provided for only for 100 scans. The LV, RV, and LVM regions are delinated and provided in the ground-truth, but the ground-truths are provided for only 100 subjects.

3.2 Performance metrics

In this work, several metrics are used to comprehensively evaluate and compare the performance of the proposed segmentation model with the other works. We used distance-based metrics, namely Hausdorff distance (HD) [28], and Mean Surface Distance (MSD) [29] to evaluate the segmentation in terms of shape accuracy of the predicted regions. These metrics are formulated as:

$$MSD = \frac{1}{2}\left(\frac{1}{n_y}\sum_{l \epsilon s_y} \min_{\mathbf{l} \epsilon \mathbf{s_y}} E_d(l,m) + \frac{1}{n_x}\sum_{m \epsilon s_x} \min_{\mathbf{m} \epsilon \mathbf{s_x}} E_d(l,m)\right) \tag{10}$$

$$HD = \max\left(\max_{l \epsilon s_y} \min_{m \epsilon s_x} E_d(l,m),\ \max_{m \epsilon s_x} \min_{l \epsilon s_y} E_d(l,m)\right) \tag{11}$$

In Equation 10, and 11, s_y, and s_x represent the surfaces of the predicted and ground truth masks with n_x, and n_y indicating the number of surface points, and the $E_d()$ represent the Euclidean distance.

In addition to distance-based metrics, we also used area-based metrics to compare the ground truth with the predicted mask. These metrics include Recall (REC) [30], Precision (PRE) [31], Jaccard Similarity Index (JSI) [32], and Dice Coefficient (DC) [33]. These metrics are formulated as follows:

$$Recall(REC) = \frac{TP_{seg}}{TP_{seg} + FN_{seg}} \tag{12}$$

$$Precision(PRE) = \frac{TP_{seg}}{TP_{seg} + FP_{seg}} \tag{13}$$

$$Dice\ Coefficient(DC) = \frac{2\,|X \cap Y|}{|X| + |Y|} \tag{14}$$

$$Jaccard\ Similarity\ Index(JSI) = \frac{|X \cap Y|}{|X \cup Y|} \tag{15}$$

In Equations (12)-(13), TP_{seg} represents the number of pixels that are correctly predicted as the region of interest (i.e, RV, LV, LVM), FP_{seg}, and FN_{seg} represents the number pixels that are wrongly predicted as the region of interest and background. In Equations (14)-(15), the X, Y represent the predicted mask, and the ground-truth.

3.3 Experimental setup

Despite the fact that 3D segmentation methods can improve the tight correlation between consecutive slices, we employed 2D slices for training and testing due to the large slice thickness and differences in slice thickness for different sequences. For making the images suitable to be fed into the proposed MMC-Net, they are resized to 224×224 pixels. The number slices are kept unchanged. The intensity of the resized images is normalized to a distribution of one variance and zero means. Due to the limited number of training samples, data augmentation is employed to increase the size of the training set. The data augmentation operation includes random rotation (-20 to -20 degrees), flipping (90 degrees), horizontal and vertical flipping. For model training, the Adam optimizer [34] is used. The learning rate is set to 0.0005, which reduces by a factor of 25 for every 20 epochs. The batch size is set to 16 and trained for 200 epochs.

The experiments are conducted using MS-CMRSeg and ACDC dataset and the results are explained in the following sections.

3.4 Performance of the proposed MMC-Net

In the first experiment, we have trained the proposed MMC-Net with all the data available in the train set of the MS-CMRSeg Challenge 2019 irrespective of the sequence. Since the submissions for the challenge are closed, we used the gold-standard T2-SPAIR, LGE, and bSSFP test set provided by the challenge organizers for evaluating the proposed MMC-Net. The MMC-Net is evaluated individually on each of the sequence, and the results achieved are tabulated in Table 1.

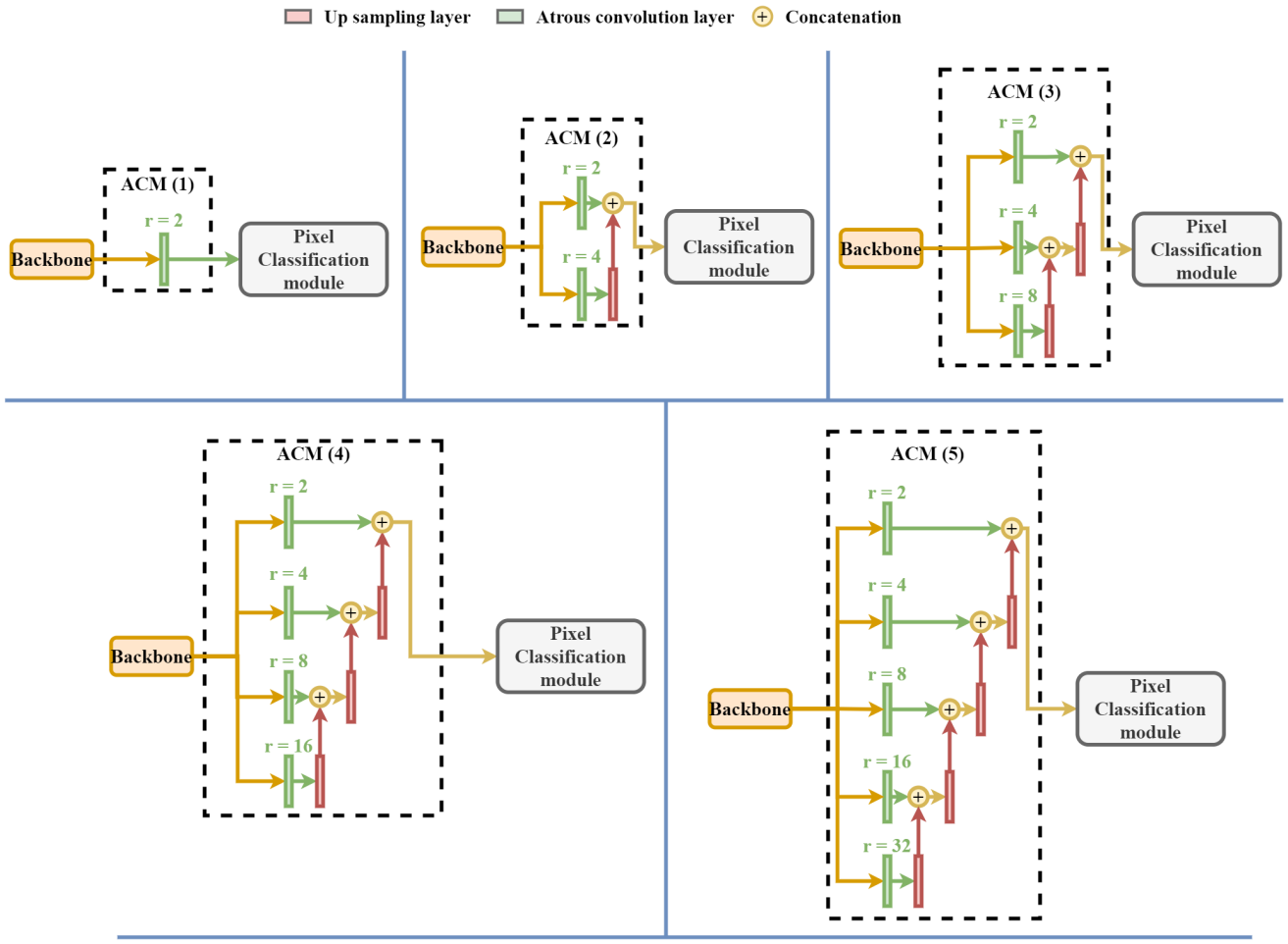

Fig. 3. ACM with different layer configurations.

TABLE 1
Performance achieved by the proposed MMC-Net.

MRI Seq.	Cardiac Struct.	HD (mm)	MSD (mm)	DC (%)	JSI (%)	REC (%)	PRE (%)
LGE	RV	7.07	0.94	93.85	94.70	93.96	94.86
	LV	7.62	0.87	96.20	94.93	96.76	97.34
	LVM	6.32	0.88	97.80	95.64	98.20	98.83
T2-SPAIR	RV	6.95	0.83	94.19	93.87	94.50	95.76
	LV	7.14	0.79	95.05	96.00	94.92	96.42
	LVM	6.83	0.81	95.86	96.11	95.90	95.97
bSSFP	RV	6.57	0.82	95.42	96.57	96.31	97.85
	LV	6.09	0.75	96.47	95.89	95.57	98.41
	LVM	6.25	0.81	96.23	96.52	95.78	97.53

As shown in Table 1, the area-based metrics, namely DC, JSI, REC, and PRE, are pretty high for all the cardiac structures. The MSD is also very low (less than 1.0 mm) for all the structures, which indicates that the predicted region boundaries are close to the ground-truth boundaries. But the HD values are a bit high, indicating the presence of minor irregularities in the segmentation results. These minor irregularities are due to the presence of noise and low contrast in the CMRI scans. The proposed MMC-Net achieved the highest performance in terms of HD, MSD, DC, JSI, REC, and PRE for LVM than the other structures. The segmentation of LVM is very difficult because they have abnormal tissues which appear similar to the blood pool, making the segmentation of LVM difficult.

3.5 Ablation study

3.5.1 *Evaluating the effectiveness of growth rate*

The growth rate, m, is a very important hyperparameter because the densely connected backbone's parametric size depends on the growth rate. In this experiment, the MMC-Net is trained and evaluated several times by varying the growth rate using the same dataset and the training strategy. The results obtained are tabulated in Table 2. The growth rate controls how much new features maps can be added to each layer. The small growth rate may not produce enough feature maps to learn the different cardiac structures. At the same time, large growth rate may shift our model towards overfitting.

As seen in Table 2, the DC, JSI, and REC of the segmentation model improves with the increase in the growth rate until the value 60. This is because, with the increase in growth rate, more convolutional filters are added to the network. Thus, enhancing the capability of the network to learn more discriminative features, thereby achieving higher DC, JSI and REC. However, when the growth rate equals to 70 the performance decreases. This indicates that the model is shifted towards overfitting. Therefore, the optimal

TABLE 2
Performance achieved with different growth rates.

Growth rate (m)	Cardiac Structure	LGE			T2-SPAIR			bSSFP		
		DC (%)	JSI (%)	REC (%)	DC (%)	JSI (%)	REC (%)	DC (%)	JSI (%)	REC (%)
20	RV	70.24	71.65	71.82	74.79	73.31	74.65	74.62	75.91	73.67
	LV	76.31	75.94	76.63	75.62	74.89	75.21	77.25	75.56	72.56
	LVM	75.09	76.27	76.38	77.51	75.35	75.42	72.46	70.15	72.67
30	RV	79.05	80.15	80.32	80.46	79.63	81.53	80.93	79.52	81.62
	LV	80.36	79.85	80.68	79.39	78.52	80.86	81.46	78.93	79.89
	LVM	80.52	81.43	81.35	80.52	81.43	79.48	80.72	80.46	81.62
40	RV	83.67	82.53	83.25	83.26	83.65	84.13	85.74	83.56	84.27
	LV	84.54	83.96	83.54	84.21	84.54	83.69	83.56	84.73	82.45
	LVM	84.89	82.94	84.62	84.96	85.23	82.46	83.84	82.51	84.24
50	RV	85.71	86.35	86.36	85.32	84.9	85.35	87.38	86.31	86.87
	LV	87.92	86.83	85.72	86.63	85.76	84.62	85.84	85.49	85.93
	LVM	87.32	87.92	88.41	86.87	85.34	84.58	84.56	85.06	85.63
60 (employed)	**RV**	**93.85**	**94.70**	**93.96**	**94.19**	**93.87**	**94.50**	**95.42**	**96.57**	**96.31**
	LV	**96.20**	**94.93**	**96.76**	**95.05**	**96.00**	**94.92**	**96.47**	**95.89**	**95.57**
	LVM	**97.80**	**95.64**	**98.20**	**95.86**	**96.11**	**95.90**	**96.23**	**96.52**	**95.78**
70	RV	89.52	88.51	87.92	89.15	90.62	88.48	90.24	87.26	89.41
	LV	90.62	88.83	88.56	88.83	87.93	88.31	88.73	87.84	86.32
	LVM	89.81	87.62	88.91	89.52	88.61	87.83	87.62	86.58	86.92

TABLE 3
Performance achieved with different ACM configurations.

Configuration	Cardiac Structure	LGE			T2-SPAIR			bSSFP		
		DC (%)	JSI (%)	REC (%)	DC (%)	JSI (%)	REC (%)	DC (%)	JSI (%)	REC (%)
Network + No ACM	RV	63.01	59.4	60.49	61.68	60.73	60.42	62.47	61.53	61.62
	LV	62.90	60.50	59.14	61.83	62.63	62.55	60.52	61.38	62.07
	LVM	63.40	66.40	60.66	60.37	61.44	61.65	63.11	62.42	63.66
Network + ACM (1)	RV	75.69	70.88	77.63	73.64	72.37	72.68	70.53	69.54	69.42
	LV	77.36	75.26	76.59	75.13	74.78	75.99	72.47	70.49	69.73
	LVM	79.46	73.46	78.22	77.42	75.07	76.94	71.33	70.32	70.28
Network + ACM (2)	RV	81.97	79.06	80.09	82.53	83.07	83.47	80.78	80.85	79.04
	LV	83.89	81.41	81.05	81.64	80.16	82.00	82.31	80.11	81.52
	LVM	85.20	79.62	84.18	83.45	82.53	82.58	83.77	83.00	82.94
Network + ACM (3)	RV	89.54	86.29	89.16	87.82	86.79	85.76	87.12	86.82	87.88
	LV	89.17	87.44	87.05	86.73	86.78	86.34	86.53	86.79	85.84
	LVM	88.62	83.35	88.53	86.84	85.09	86.41	86.25	87.35	87.43
Network + ACM (4) (proposed)	**RV**	**93.85**	**94.70**	**93.96**	**94.19**	**93.87**	**94.50**	**95.42**	**96.57**	**96.31**
	LV	**96.20**	**94.93**	**96.76**	**95.05**	**96.00**	**94.92**	**96.47**	**95.89**	**95.57**
	LVM	**97.80**	**95.64**	**98.20**	**95.86**	**96.11**	**95.90**	**96.23**	**96.52**	**95.78**
Network + ACM (5)	RV	88.29	84.21	87.14	90.63	88.35	88.56	87.65	88.52	87.94
	LV	87.94	85.00	86.94	89.32	89.54	88.93	89.64	87.93	86.03
	LVM	86.74	80.78	86.31	89.73	88.63	89.16	88.62	87.64	86.84

segmentation performance is obtained when the growth rate is set to 60.

3.5.2 Evaluating the effectiveness of ACM

For understanding the effectiveness of the proposed ACM module, we first removed the ACM from the network and evaluated the performance. Then, we added the ACM to the network and tried with different layer configurations (one-layered to five layered). The different ACM configurations are shown in Fig. 3. The DC, JSI, and REC achieved by these networks on the gold-standard test set is shown in Table 3. It shows that the networks with ACM achieved better DC, JSI, and REC than the network with no ACM. This shows the importance of ACM for enhancing the performance in cardiac image segmentation.

Among the different layered configurations of ACM, the ACM with four atrous convolutions achieved better DC, JSI, and REC in the segmentation of LV, RV, and LVM than the other configurations. This demonstrates that the receptive field captured by the four-layered configuration is suitable for learning more discriminative features of LV, RV, and LVM than the one-layered, two-layered, three-layered, and five-layered configurations. The receptive fields captured by one-layered, two-layered, and three-layered configurations are not big enough to learn the features of the cardiac structures, whereas the receptive field captured by the five-layered configuration is too large to mix-up with the background.

3.5.3 Evaluating the effectiveness of upsampling technique used in ACM

The ACM module contains four atrous convolution layers arranged hierarchically in this work, as shown in Fig. 2. In the ACM module, the output feature map of the lower atrous convolution layer is upsampled to be concatenated with the output of the above atrous convolution layer. There are two well-known upsampling techniques, namely de-convolution, and bi-linear interpolation. To understand the

performance affected by these two upsampling techniques, we have trained and evaluated the MMC-Net individually with these two up-sampling techniques. The DC, JSI, and REC achieved with different upsampling techniques are summarised in Table 4.

The experimental results tabulated in Table 4 shows that the model using the deconvolution technique in the ACM achieved better JSI, DC, and REC than the model with bilinear interpolation. This is because the deconvolution is trained during the training phase, which means that the filter's weights can adjust themselves to an appropriate value for optimizing the objective function. In bilinear interpolation, the output feature maps are upsampled three times their original size and then cropped to match the network's receptive field. As a result, some redundant information will be introduced, impacting the overall segmentation performance. Therefore, deconvolution is more suitable for this work than bilinear interpolation.

3.6 Comparison with State-of-the-art Models

The performance achieved by the proposed segmentation model is compared with the state-of-the-art segmentation methods, including DUpsample [35], UNet++ [36], SegNet [37], and DeepLabV3+ [38]. All these methods employed an encoder-decoder network structure. SegNet is one of the earlier segmentation works proposed for multi-class segmentation tasks. The DeepLabV3+ is another segmentation network that has achieved state-of-the-art performance on Cityscapes and PASCAL VOC 2012 datasets. UNet++ is an extended version of UNet, where a set of dense and nested skip paths are added to the UNet. This network has shown impressive performance on several medical image datasets, including poly segmentation [36], nuclei segmentation, and lung nodule segmentation. The DUpsample is the most recent method proposed for enhancing the upsampling operation in the decoder for recovering pixel-wise predictions from low-resolution images. This network has shown state-of-the-art performance on the PASCAL VOC 2012 dataset. For a fair comparison, these models are downloaded from their public implementations and trained and evaluated with the same evaluation protocol and parameter setting as the proposed MMC-Net.

3.6.1 LGE CMRI

The performance comparison is tabulated in Table 5, and sample segmentation results on apical, middle, and basal slices are shown in Fig. 4. As presented in Table 5 and Fig. 4, DeepLabV3+ has achieved worse performance than the other methods in cardiac segmentation. The other methods achieved better performance than the DeepLabV3+, but they faced difficulty in segmenting the LVM from apical slices. However, manual delineation of LVM is also very hard for experienced cardiologists, radiologists, and other medical professionals. Because the boundaries of LVM are not clear in the apical slice. The middle slice has better contrast, so the boundaries of the LV, RV, and LVM are much clearer. As shown in Fig. 4, all models except the DeepLabV3+ segmented the myocardium (LVM) and ventricular structures (LV and RV) better in middle slice than in the other slices. However, the segmentation performance achieved by the SegNet is not so accurate; this may be due to the lack of skip connections between the decoder and the encoder. The DUpsample achieved better performance than the UNet, SegNet, and DeepLabV3+, but it over-segmented the LVM and RV on the basal and middle slices (Fig. 4). Overall, the proposed segmentation method achieved better recall than the other models, and predicted results are mostly matched with the ground truth, as shown in Fig. 4.

3.6.2 bSSFP CMRI

The quantitative performance achieved by the proposed MMC-Net and the state-of-the-art models is presented in Table 6. It can be seen that the proposed segmentation model achieved higher performance than the other state-of-the-art methods on the gold-standard test set. The segmentation results produced by these models on sample bSSFP CMR slices is presented in Fig 5. Except the DeepLabV3+, all other models obtained better segmentation results for LV, and RV on all the three slices. The DeepLabV3+ poorly delinated the LV, RV, and LVM in all the three slices. Only the MMC-Net obtained accurate segmentation results on all the three slices. The UNet++ and the SegNet incorrectly segmented the background for LVM in the basal slice as shown in Fig. 5.

3.6.3 T2-SPAIR CMRI

The comparative results reported by the proposed MMC-Net and the state-of-the-art models are presented in Table 7. The pictorial comparisons is shown in Fig. 6. The MMC-Net achieved highest DC, and JSI than the other state-of-the-art-methods. In T2-SPAIR CMR images, the DeepLabV3+ performed better than on the LGE, and bSSFP slices. The UNet++ reported poor results than the other state-of-the-art models. The segmentation of RV by DUpsample, and SegNet on the basal slice is not so accurate. Compared to DUpsample, and SegNet, the DeepLabV3+ produced better segmentation of RV. But the DeepLabV3+ produced better segmentation result for LV on the basal slice. Our MMC-Net reported accurate segmentation results for LV, RV, and LVM in all the three slices.

3.7 Comparison with the existing studies in the literature

In this work, we have compared the performance achieved by the proposed MMC-Net with three existing methods namely Li et al. [39], SK-Unet [40], and Li et al. [41]. Among these approaches, the Li et al. [39], and Li et al. [41] experimented on all the three modalities in the dataset (i.e, bSSFP, LGE, and T2-SPAIR), but SK-Unet [40] was experimented on only LGE CMRI. These approaches were experimented with different evaluation protocol like Li et al. [39] used 4-fold cross validation, Li et al. [41] used 5-fold cross-validation, and SK-Unet [40] used the entire training set provided in the MS-CMRSeg-2019 dataset for training, and tested only on the LGE CMRI in the test set. So for a fair comparison, we have also evaluated our MMC-Net using 4-fold, and 5-fold cross evaluation protocol. Since the evaluation protocol employed in this work (i.e, as presented in sub-section 3.4) is as the one used by SK-Unet [40], we have compared directly without reevaluation. This comparison is tabulated

TABLE 4
Performance achieved by Bilinear interpolation and deconvolution upsampling techniques.

Upsampling technique	Cardiac Structure	LGE			T2-SPAIR			bSSFP		
		DC (%)	JSI (%)	REC (%)	DC (%)	JSI (%)	REC (%)	DC (%)	JSI (%)	REC (%)
Bi-linear interpolation	RV	87.26	85.72	82.56	86.83	85.72	86.03	84.13	83.37	84.62
	LV	88.41	86.00	87.62	83.51	84.36	82.22	85.06	84.86	84.45
	LVM	82.53	85.53	86.04	84.24	84.68	84.72	84.53	84.07	84.92
Deconvolution (employed)	**RV**	**93.85**	**94.70**	**93.96**	**94.19**	**93.87**	**94.50**	**95.42**	**96.57**	**96.31**
	LV	**96.20**	**94.93**	**96.76**	**95.05**	**96.00**	**94.92**	**96.47**	**95.89**	**95.57**
	LVM	**97.80**	**95.64**	**98.20**	**95.86**	**96.11**	**95.90**	**96.23**	**96.52**	**95.78**

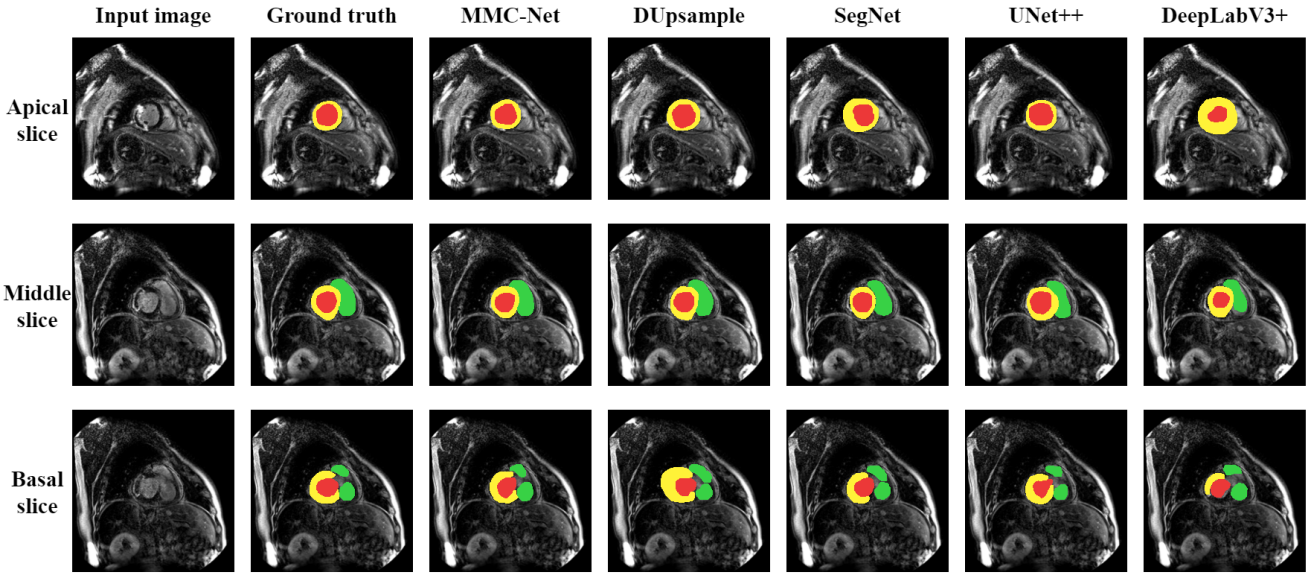

Fig. 4. Comparison of segmentation performance on sample LGE CMRI slices; the red, green, and yellow regions represents the LV, RV, and LVM.

TABLE 5
Performance comparison with the state-of-the-art-networks in segmenting LGE CMRI.

Methods	Cardiac structures	DC (%)	JSI (%)	REC (%)
DUpsample [35]	RV	85.60	85.80	89.83
	LV	88.70	80.30	93.45
	LVM	87.30	77.80	88.10
UNet++ [36]	RV	83.90	84.20	82.53
	LV	85.20	78.40	85.96
	LVM	84.70	72.90	86.73
SegNet [37]	RV	76.00	83.90	82.27
	LV	84.40	77.40	80.95
	LVM	83.10	70.70	79.07
DeepLabV3+ [38]	RV	59.10	64.40	61.32
	LV	59.60	59.50	61.90
	LVM	61.80	56.40	62.33
MMC-Net (Proposed)	**RV**	**93.85**	**94.70**	**93.96**
	LV	**96.20**	**94.93**	**96.76**
	LVM	**97.80**	**95.64**	**98.20**

TABLE 6
Performance comparison with the state-of-the-art-networks in segmenting bSSFP CMRI.

Methods	Cardiac structures	DC (%)	JSI (%)	REC (%)
DUpsample [35]	RV	86.42	85.92	83.89
	LV	81.85	84.08	84.16
	LVM	86.62	85.00	83.68
UNet++ [36]	RV	78.69	79.56	77.93
	LV	80.15	81.42	73.78
	LVM	77.05	77.63	78.64
SegNet [37]	RV	77.72	75.79	74.96
	LV	79.86	78.31	75.70
	LVM	76.35	75.93	73.07
DeepLabV3+ [38]	RV	57.91	64.40	61.32
	LV	59.62	59.50	61.90
	LVM	58.36	56.40	62.33
MMC-Net (Proposed)	**RV**	**95.42**	**96.57**	**96.31**
	LV	**96.47**	**95.89**	**95.57**
	LVM	**96.23**	**96.52**	**95.78**

in Table 8. The proposed MMC-Net outperformed all the three existing approaches. The performance enhancement achieved by the MMC-Net on all the three MRI modalities is phenomenal. Especially compared to SK-Unet [40], the DC enhancement is 9.1% in segmenting LVM for LGE CMRI. Also, compared to Li et al. [39], the enhancement is 6.21% in segmenting LV for T2-SPAIR CMRI, and compared to Li et al. [41], the DC enhancement is 17.52% in segmenting RV for bSSFP CMRI.

3.8 Generalization ability of the proposed MMC-Net

We further evaluate the proposed MMC-Net on another publicly available challenging dataset the ACDC [27]. For understanding the generalization ability, we trained and evaluated our model on the ACDC dataset without any fine-tuning based on the experimental setup provided in [42].

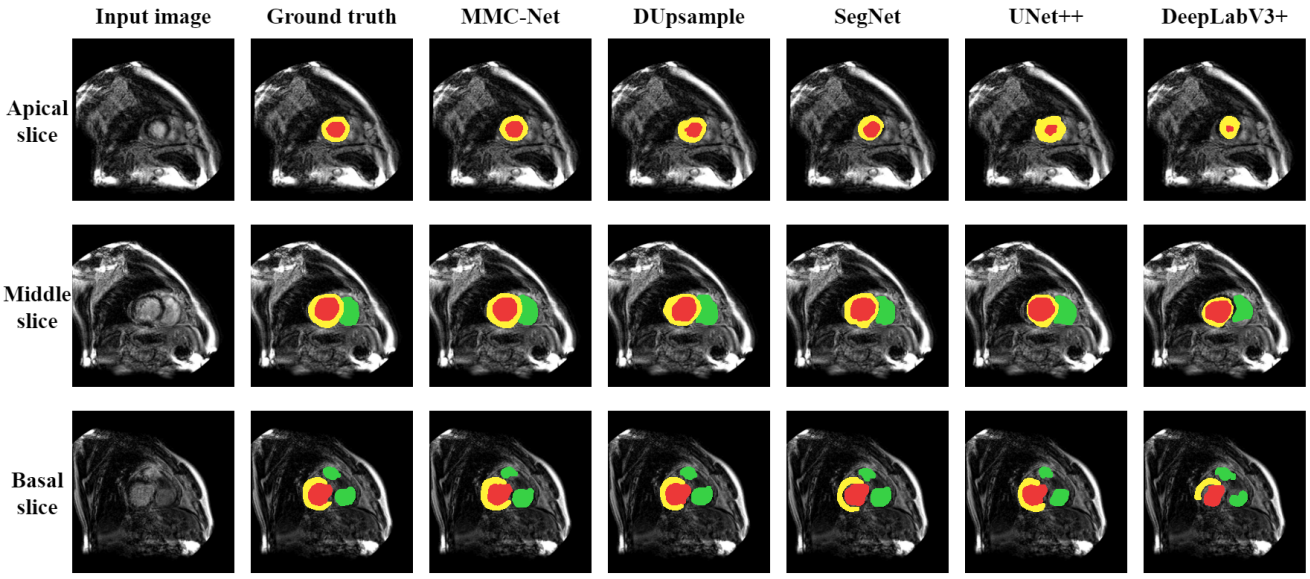

Fig. 5. Comparison of segmentation performance on sample bSSFP CMRI slices; the red, green, and yellow regions represents the LV, RV, and LVM.

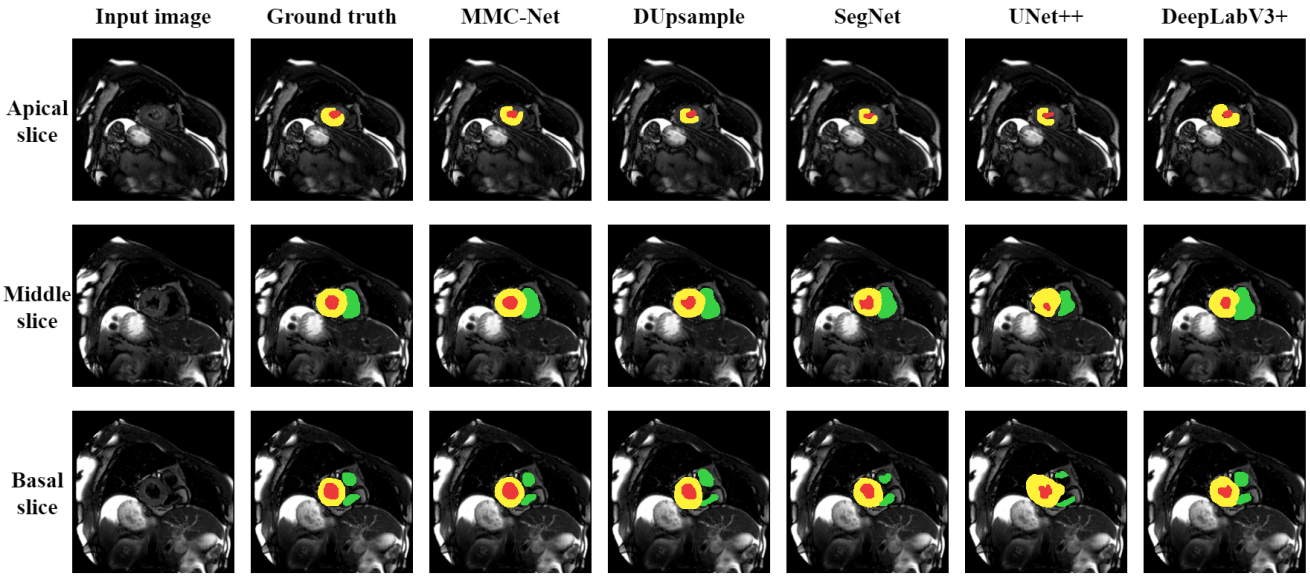

Fig. 6. Comparison of segmentation performance on sample T2-SPAIR CMRI slices; the red, green, and yellow regions represents the LV, RV, and LVM.

The results achieved by the MMC-Net on the ACDC dataset are discussed below:

3.8.1 *Comparison with state-of-the-art*

We compared the segmentation results produced by the MMC-Net with the state-of-the-art methods employing the ACDC dataset. This comparison is tabulated in Table 9. As seen in Table 9, the proposed MMC-Net achieved better performance than the state-of-the-art methods. Except for the UNet++, all other state-of-the-art methods achieved better HD, and DC in both ED, and ES phases. The SegNet achieved the highest HD, and DC among the state-of-the-art methods.

The qualitative segmentation results produced by these methods are presented in Fig 7. From Fig 7, it can be seen that the segmentation mask predicted by the MMC-Net is close to the ground truth. The UNet++ produced under-segmentation results on all three slices. The DUpsample slightly over segmented the RV in the middle slice, whereas the DeepLabV3+ slightly over segmented the RV, LVM in the middle, and the basal slices.

3.8.2 *Comparison with the methods listed on competition leader board.*

We also compared the performance achieved by the proposed MMC-Net with methods listed on the ACDC challenge leaderboard. These results are tabulated in Table 10. Only the top-four methods [19], [43], [44], [45] from challenge leader board are used for comparison. We briefly discuss the architectures of the above mentioned methods:

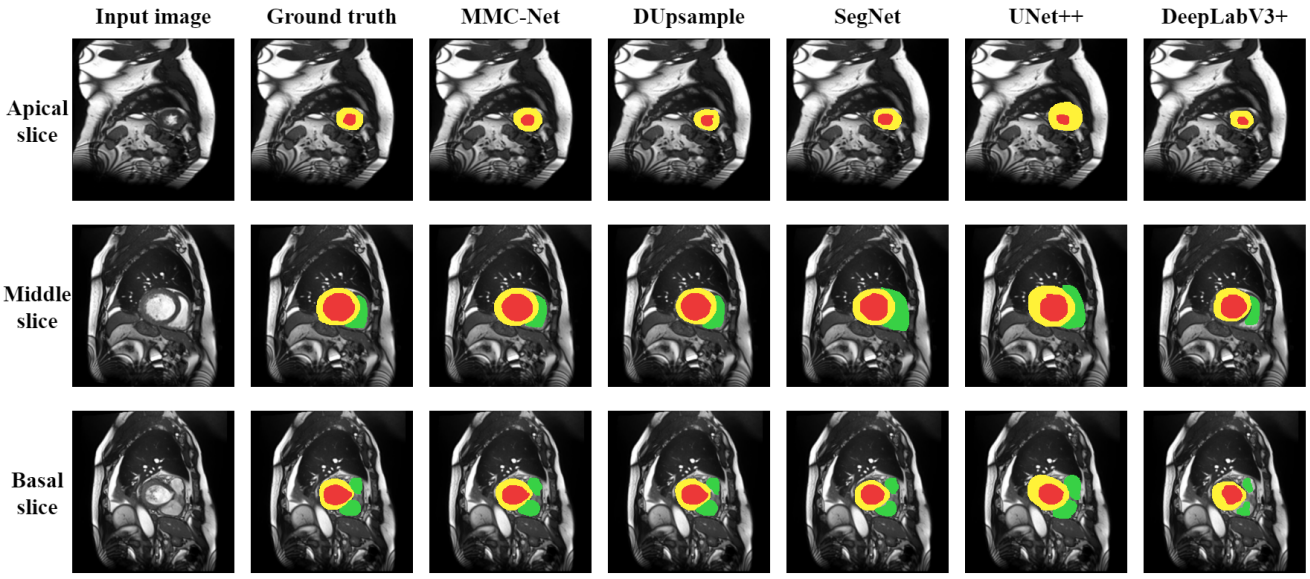

Fig. 7. Comparison of segmentation performance on sample slices from ACDC dataset; the red, green, and yellow regions represents the LV, RV, and LVM.

TABLE 7
Performance comparison with the state-of-the-art-networks in segmenting T2-SPAIR CMRI.

Methods	Cardiac structures	DC (%)	JSI (%)	REC (%)
DUpsample [35]	RV	85.69	87.42	87.42
	LV	85.78	88.37	86.94
	LVM	86.25	87.91	87.72
UNet++ [36]	RV	68.42	70.42	69.14
	LV	70.12	71.57	70.62
	LVM	70.34	71.83	71.58
SegNet [37]	RV	83.47	82.64	83.75
	LV	84.92	83.52	85.42
	LVM	85.78	85.89	84.67
DeepLabV3+ [38]	RV	72.68	73.36	72.62
	LV	73.72	74.19	73.85
	LVM	74.51	73.96	72.26
MMC-Net (Proposed)	**RV**	**94.19**	**93.87**	**94.50**
	LV	**95.05**	**96.00**	**94.92**
	LVM	**95.86**	**96.11**	**95.90**

Zotti et al. [45] employed a multi-resolution gradient structure, which can be considered to be an extension to the U-Net, Khened et al. [43] developed a two stage U-Net based network by embedding Densely connected blocks in the place of convolutional layers in the standard U-Net, Isensee et al. [44] developed an ensemble of 10 models based on 2D, and 3D UNet architectures, and Painchaud et al. [19] designed an adhersal variational autoencoder for producing better segmentation results.

As shown in Table 10, the proposed MMC-Net has achieved better DC than the other four models in segmenting all the three structures. Among the other methods, the 2D-3D U-Net ensemble approach proposed by the Isense [44] achived highest performance. When compared to 2D-3D U-Net, the proposed MMC-Net enhanced the DC in segmenting LV via ED and ES phases by 0.3% and 2.3% respectively. Also, compared to 2D-3D U-Net, the proposed MMC-Net enhanced the DC in segmenting RV and LVM via ED and ES phases by 2.4%, 5.7%, 4.3%, and 4.1% respectively.

3.8.3 *Comparison with existing approaches*

We have also compared the segmentation DC, HD achieved by the proposed MMC-Net with the three recently proposed methods [46], [47], [48] that were not listed on the leader board. This comparison is tabulated in Table 11. As shown in Table 11, the proposed method outperformed the methods listed in the table. Overall, we conclude that the proposed MMC-Net is a more efficient method for segmenting LV, RV and LVM than other existing methods in the literature. This shows a powerful generalization performance of the proposed MMC-Net.

4 Conclusion

This work presents a novel convolutional neural network model named Multi-Modal Cardiac Net (MMC-Net) for the automatic segmentation of RV, LV, and LVM from multi-modal CMRI scans. In this network, a densely connected backbone enables feature reuse for retaining more useful information. And we propose an atrous-convolution module to integrate with the backbone for multi-scale context-learning. The experimental results illustrate that the proposed segmentation model achieved higher performance than the state-of-the-art methods and the other studies found in the literature. The proposed MMC-Net's significant achievement is that with just five LGE scans and 35 T2-SPAIR and bSSFP scans from the MS-CMRSeg for training, the proposed MMC-Net was able to segment RV, LV, and LVM from the LGE, T2-SPAIR, and bSSFP CMR scans with high precision and recall.

In addition, based on the experiment results, the proposed MMC-Net outperformed all the advanced methods without any fine-tuning on the ACDC dataset. This shows a powerful generalization performance of the MMC-Net. In the future, the studies can focus on training and evaluating their models on data from various modalities, MRI

TABLE 8
Performance comparison with existing approaches employing MS-CMRSeg 2019 dataset.

Methods	Evaluation protocol	LGE			T2-SPAIR			bSSFP		
		RV	LV	LVM	RV	LV	LVM	RV	LV	LVM
Li et al. [39]	4-Fold	98.68	92.29	96.73	96.77	91.94	95.69	91.52	94.93	96.08
Li et al. [41]	5-Fold	83.20	91.90	96.5	81.30	92.40	85.70	77.90	94.60	83.10
SK-Unet [40]	Official split	93.20	95.80	88.70	-	-	-	-	-	-
MMC model	4-Fold	**99.12**	**96.69**	**97.64**	**97.19**	**98.15**	**98.49**	**97.37**	**97.79**	**98.17**
	5-Fold	**98.36**	**97.89**	**98.51**	**98.53**	**97.31**	**96.73**	**97.69**	**97.36**	**97.05**
	Official split	**93.85**	**96.20**	**97.80**	94.19	95.05	95.86	95.42	96.47	96.23

TABLE 9
Comparison of state-of-the-art performance on ACDC dataset.

Methods	LV				RV				LVM			
	HD		DC		HD		DC		HD		DC	
	ED	ES	ED	ES	ED	ES	ED	ES	ED	ES	ED	ES
DUpsample [35]	7.34	7.40	89.13	90.67	8.10	7.80	89.68	87.82	7.40	7.29	89.52	90.46
UNet++ [36]	8.39	8.54	79.62	80.35	9.47	9.63	80.18	79.92	8.94	8.91	75.25	74.95
SegNet [37]	7.10	7.30	89.61	91.23	7.90	7.70	90.52	88.79	6.80	7.00	91.83	92.93
DeepLabV3+ [38]	7.59	8.01	86.69	84.94	8.42	8.00	87.91	83.61	7.82	7.55	86.26	85.91
MMC-Net (Proposed)	**5.90**	**6.50**	**97.10**	**95.40**	**6.70**	**5.30**	**97.00**	**95.60**	**5.40**	**6.20**	**94.50**	**96.00**

TABLE 10
Comparison of DC, and HD with methods presented on the ACDC challenge.

Methods	LV				RV				LVM			
	HD		DC		HD		DC		HD		DC	
	ED	ES	ED	ES	ED	ES	ED	ES	ED	ES	ED	ES
Isensee et al. [44]	7.38	6.90	96.80	93.10	10.12	12.14	94.60	89.90	8.72	8.67	90.20	91.90
Zotti et al. [45]	6.64	8.70	95.70	90.50	10.31	14.05	94.10	88.20	9.60	9.30	89.00	90.00
Painchaud et al. [19]	6.10	8.30	96.10	91.10	13.70	13.30	93.30	88.40	8.60	9.60	88.10	89.70
Khened et al. [43]	8.12	8.96	96.40	91.70	13.99	13.93	93.50	87.90	9.84	12.58	88.90	89.80
MMC-Net (Proposed)	**5.90**	**6.50**	**97.10**	**95.40**	**6.70**	**5.30**	**97.00**	**95.60**	**5.40**	**6.20**	**94.50**	**96.00**

TABLE 11
Comparison of DC, and HD with the recently proposed methods.

Methods	LV				RV				LVM			
	HD		DC		HD		DC		HD		DC	
	ED	ES	ED	ES	ED	ES	ED	ES	ED	ES	ED	ES
DBAN [46]	6.70	8.10	96.00	90.00	10.60	12.60	94.00	89.00	8.80	8.70	85.00	88.00
Silva et al. [47]	8.06	10.40	96.30	91.20	14.60	17.50	90.00	86.00	7.90	9.90	89.40	90.50
Zhang et al. [48]	6.26	8.42	96.30	91.50	13.06	13.93	92.20	87.20	8.65	9.53	89.00	90.10
MMC-Net (Proposed)	**5.90**	**6.50**	**97.10**	**95.40**	**6.70**	**5.30**	**97.00**	**95.60**	**5.40**	**6.20**	**94.50**	**96.00**

manufacturers, and medical centers across the globe. And include statistical shape models for improving the precision in locating boundaries of cardiac structures.

REFERENCES

[1] B. Ruijsink, E. Puyol-Antón, I. Oksuz, M. Sinclair, W. Bai, J. A. Schnabel, R. Razavi, and A. P. King, "Fully automated, quality-controlled cardiac analysis from cmr: validation and large-scale application to characterize cardiac function," *Cardiovascular Imaging*, vol. 13, no. 3, pp. 684–695, 2020.

[2] X. Zhuang, "Multivariate mixture model for cardiac segmentation from multi-sequence mri," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 581–588.

[3] Z. Xiahai, "Multivariate mixture model for myocardial segmentation combining multi source images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 12, pp. 2933–2946, 2019.

[4] M. Khened, V. A. Kollerathu, and G. Krishnamurthi, "Fully convolutional multi-scale residual densenets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers," *Medical image analysis*, vol. 51, pp. 21–45, 2019.

[5] L. Grosse-Wortmann, C. K. Macgowan, L. Vidarsson, and S.-J. Yoo, "Late gadolinium enhancement of the right ventricular myocardium: is it really different from the left?" *Journal of Cardiovascular Magnetic Resonance*, vol. 10, no. 1, pp. 1–9, 2008.

[6] M. Vöhringer, H. Mahrholdt, A. Yilmaz, and U. Sechtem, "Significance of late gadolinium enhancement in cardiovascular magnetic resonance imaging (cmr)," *Herz Kardiovaskuläre Erkrankungen*, vol. 32, no. 2, pp. 129–137, 2007.

[7] M. A. Fagiry, I. Abdelaziz, R. Davidson, and M. Z. Mahmoud, "The recent advances, drawbacks, and the future directions of cmri in the diagnosis of ihd," *Scientific Reports*, vol. 11, no. 1, pp. 1–9, 2021.

[8] C. Ciofolo, M. Fradkin, B. Mory, G. Hautvast, and M. Breeuwer, "Automatic myocardium segmentation in late-enhancement mri," in *2008 5th IEEE International Symposium on Biomedical Imaging: from nano to macro*. IEEE, 2008, pp. 225–228.

[9] E. Dikici, T. O'Donnell, R. Setser, and R. D. White, "Quantification of delayed enhancement mr images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2004, pp. 250–257.

[10] R. El Berbari, N. Kachenoura, F. Frouin, A. Herment, E. Mousseaux, and I. Bloch, "An automated quantification of the transmural myocardial infarct extent using cardiac de-mr images," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2009, pp. 4403–4406.

[11] D. Wei, Y. Sun, S.-H. Ong, P. Chai, L. L. Teo, and A. F. Low, "Three-dimensional segmentation of the left ventricle in late gadolinium enhanced mr images of chronic infarction combining long-and short-axis information," *Medical image analysis*, vol. 17, no. 6, pp. 685–697, 2013.

[12] R. S. Xu, P. Athavale, Y. Lu, P. Radau, and G. A. Wright, "Myocardial segmentation in late-enhancement mr images via registration and propagation of cine contours," in *2013 IEEE 10th International Symposium on Biomedical Imaging*. IEEE, 2013, pp. 856–859.

[13] Q. Tao, S. R. Piers, H. J. Lamb, and R. J. van der Geest, "Automated left ventricle segmentation in late gadolinium-enhanced mri for objective myocardial scar assessment," *Journal of Magnetic Resonance Imaging*, vol. 42, no. 2, pp. 390–399, 2015.

[14] Y. Lu, G. Wright, and P. E. Radau, "Automatic myocardium segmentation of lge mri by deformable models with prior shape data," *Journal of Cardiovascular Magnetic Resonance*, vol. 15, no. 1, pp. 1–2, 2013.

[15] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 4, pp. 640–651, 2016.

[16] P. V. Tran, "A fully convolutional neural network for cardiac segmentation in short-axis mri," *arXiv preprint arXiv:1604.00494*, 2016.

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[18] J. Li, Z. L. Yu, Z. Gu, H. Liu, and Y. Li, "Dilated-inception net: multi-scale feature aggregation for cardiac right ventricle segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 12, pp. 3499–3508, 2019.

[19] N. Painchaud, Y. Skandarani, T. Judge, O. Bernard, A. Lalande, and P.-M. Jodoin, "Cardiac segmentation with strong anatomical guarantees," *IEEE transactions on medical imaging*, vol. 39, no. 11, pp. 3703–3713, 2020.

[20] I. Oksuz, J. R. Clough, B. Ruijsink, E. P. Anton, A. Bustin, G. Cruz, C. Prieto, A. P. King, and J. A. Schnabel, "Deep learning-based detection and correction of cardiac mr motion artefacts during reconstruction for high-quality segmentation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 12, pp. 4001–4010, 2020.

[21] R. P. Poudel, P. Lamata, and G. Montana, "Recurrent fully convolutional neural networks for multi-slice mri cardiac segmentation," in *Reconstruction, segmentation, and analysis of medical images*. Springer, 2016, pp. 83–94.

[22] D. Zhang, I. Icke, B. Dogdas, S. Parimal, S. Sampath, J. Forbes, A. Bagchi, C.-L. Chin, and A. Chen, "A multi-level convolutional lstm model for the segmentation of left ventricle myocardium in infarcted porcine cine mr images," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 470–473.

[23] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[24] J.-M. Combes, A. Grossmann, and P. Tchamitchian, *Wavelets: Time-Frequency Methods and Phase Space Proceedings of the International Conference, Marseille, France, December 14–18, 1987*. Springer Science & Business Media, 2012.

[25] M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian, "A real-time algorithm for signal analysis with the help of the wavelet transform," in *Wavelets*. Springer, 1990, pp. 286–297.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[27] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P.-A. Heng, I. Cetin, K. Lekadir, O. Camara, M. A. G. Ballester *et al.*, "Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved?" *IEEE transactions on medical imaging*, vol. 37, no. 11, pp. 2514–2525, 2018.

[28] J. Henrikson, "Completeness and total boundedness of the hausdorff metric," *MIT Undergraduate Journal of Mathematics*, vol. 1, no. 69-80, p. 10, 1999.

[29] V. Yeghiazaryan and I. D. Voiculescu, "Family of boundary overlap metrics for the evaluation of medical image segmentation," *Journal of Medical Imaging*, vol. 5, no. 1, p. 015006, 2018.

[30] R. Trevethan, "Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice," *Frontiers in public health*, vol. 5, p. 307, 2017.

[31] F. C. Monteiro and A. C. Campilho, "Performance evaluation of image segmentation," in *International Conference Image Analysis and Recognition*. Springer, 2006, pp. 248–259.

[32] P. Jaccard, "The distribution of the flora in the alpine zone. 1," *New phytologist*, vol. 11, no. 2, pp. 37–50, 1912.

[33] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[34] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[35] Z. Tian, T. He, C. Shen, and Y. Yan, "Decoders matter for semantic segmentation: Data-dependent decoding enables flexible feature aggregation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3126–3135.

[36] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep learning in medical image analysis and multimodal learning for clinical decision support*. Springer, 2018, pp. 3–11.

[37] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[38] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.

[39] W. Li, L. Wang, F. Li, S. Qin, and B. Xiao, "Myocardial pathology segmentation of multi-modal cardiac mr images with a simple but efficient siamese u-shaped network," *Biomedical Signal Processing and Control*, vol. 71, p. 103174, 2022.

[40] X. Wang, S. Yang, Y. Fang, Y. Wei, M. Wang, J. Zhang, and X. Han, "Sk-unet: An improved u-net model with selective kernel for the segmentation of lge cardiac mr images," *IEEE Sensors Journal*, vol. 21, no. 10, pp. 11 643–11 653, 2021.

[41] F. Li, W. Li, S. Qin, and L. Wang, "Mdfa-net: Multiscale dual-path feature aggregation network for cardiac segmentation on multi-sequence cardiac mr," *Knowledge-Based Systems*, vol. 215, p. 106776, 2021.

[42] J. Pi, Y. Qi, M. Lou, X. Li, Y. Wang, C. Xu, and Y. Ma, "Fs-unet: Mass segmentation in mammograms using an encoder-decoder architecture with feature strengthening," *Computers in Biology and Medicine*, vol. 137, p. 104800, 2021.

[43] M. Khened, V. Alex, and G. Krishnamurthi, "Densely connected fully convolutional network for short-axis cardiac cine mr image segmentation and heart diagnosis using random forest," in *International Workshop on Statistical Atlases and Computational Models of the Heart*. Springer, 2017, pp. 140–151.

[44] F. Isensee, P. F. Jaeger, P. M. Full, I. Wolf, S. Engelhardt, and K. H. Maier-Hein, "Automatic cardiac disease assessment on cine-mri via time-series segmentation and domain specific features," in *International workshop on statistical atlases and computational models of the heart*. Springer, 2017, pp. 120–129.

[45] C. Zotti, Z. Luo, A. Lalande, and P.-M. Jodoin, "Convolutional neural network with shape prior applied to cardiac mri segmentation," *IEEE journal of biomedical and health informatics*, vol. 23, no. 3, pp. 1119–1128, 2018.

[46] X. Yang, Y. Zhang, B. Lo, D. Wu, H. Liao, and Y.-T. Zhang, "Dban: adversarial network with multi-scale features for cardiac mri segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 6, pp. 2018–2028, 2020.

[47] I. F. S. da Silva, A. C. Silva, A. C. de Paiva, and M. Gattass, "A cascade approach for automatic segmentation of cardiac structures in short-axis cine-mr images using deep neural networks," *Expert Systems with Applications*, p. 116704, 2022.

[48] H. Zhang, W. Zhang, W. Shen, N. Li, Y. Chen, S. Li, B. Chen, S. Guo, and Y. Wang, "Automatic segmentation of the cardiac mr images based on nested fully convolutional dense network with dilated convolution," *Biomedical signal processing and control*, vol. 68, p. 102684, 2021.

G. Jignesh Chowdary was born in 2000. He is currently pursing his Bachelor degree in computer science at the Vellore Institute of Technology, Chennai. His research interest include Medical Image Analysis, and Cyber-Physical Systems in Healthcare.

Pratheepan Yogarajah received the firstclass honors degree in computer science from the University of Jaffna, Jaffna, Sri Lanka, in 2001, the M.Phil. degree in computer vision from Oxford Brookes University, Oxford, U.K., in 2006, and the Ph.D. degree in computing and engineering from Ulster University, Derry∼Londonderry, U.K., in 2015. Currently, he is a Lecturer in computer science in the School of Computing and Intelligent Systems, Ulster University. His research interests include biometrics, computer vision, medical image analysis, steganography and digital watermarking, robotics, and machine learning. Dr. Yogarajah received the Oxford Brookes University HMGCC Scholarship Award in 2005, and coreceived the Proof of Principle Award from Ulster University in 2012 and the Proof of Concept from Invest Northern Ireland in 2013.

Priyanka Chaurasia received a BTech degree in information technology from the Harcourt Butler Technical University, Kanpur, India, in 2006 and a PhD in computing and information engineering from Ulster University, Coleraine, UK, in 2013. Currently, she is a Lecturer in Data Analytics at the School of Computing Intelligent Systems, Ulster University, Londonderry, UK. Before her Ph.D., she was with IBM India Software Labs, Bangalore, India, for 3 years as a software engineer. She has patents granted by the U.S. Patent and Trademark Office in the area of signature verification and is a recipient of the IBM Invention Achievement Award in 2008. Her research interests are assistive technology, healthcare, activity recognition, biometric security, and data analytics.