How to Cite:
Gomathi, G., & Subha, V. (2022). DPA-UNet: Detail preserving attention UNet for cardiac MRI ventricle region segmentation. *International Journal of Health Sciences*, *6*(S1), 11833–11852. https://doi.org/10.53730/ijhs.v6nS1.7903

DPA-UNet: Detail preserving attention UNet for cardiac MRI ventricle region segmentation

G. Gomathi
Research Scholar, Department of Computer Science and Engineering, Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu, India
Email: gomathi1986@gmail.com

Dr. V. Subha
Assistant Professor, Department of Computer Science and Engineering, Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu, India
Email: subha_velappan@yahoo.co.in

***Abstract*---**Cardiac Magnetic Resonance Image (MRI) plays an integral part in examining the clinical problems of cardiac disorders. The purpose is to extract effectual information from ventricle regions in which the structure varies due to systolic and diastolic phases of the heart. The proposed methodology – Detail Preserving Attention UNet (DPA-UNet) improves the level of clinical diagnosis to identify the requisite parts of analysis precisely. Thus it provides the best possible results with appropriate accuracy. In recent years, the attention and awareness towards deep learning approaches have widely been increased, where a neural network can automatically learn image features. The latter is in direct disparity with conventional deep learning approaches. UNet with attention-based models is considered the most crucial semantic segmentation framework for Cardiac MRI. The ventricle regions can be extracted by the methodology of DPA-UNet which performs through two modules such as spatial and channel attention. Initially, the channel attention module is processed with max-pool to extract the dominant features to segment the area of concentration. Subsequently, the spatial attention module is processed with an average pool to extract the global features of the ventricular regions. The process of integration proceeds subsequently. The results of the experiment show that DPA-UNet achieves the highest performance according to the metrics Dice Coefficient and Hausdorff Distance.

***Keywords*---**Cardiac MRI, DPA-UNet, Semantic Segmentation, Channel Attention, Spatial Attention.

Introduction and Background Study

This section was further subdivided into two sections. Section 1.1 gives an introduction to cardiovascular diseases and a general introduction to the methodologies used in this paper. Section 1.2 presents an overview of the methodologies that were used by existing authors in different areas. Finally, this section was concluded with the author's contributions.

Introduction

Cardiovascular diseases (CVDs) refer to circumstances that affect the structures and functions of the human heart. World health organization (WHO) says that CVDs were considered as the foremost causes of death worldwide, taking around 17.9 million lives every year. Recent medical imaging approaches like Magnetic Resonance Imaging (MRI), Computed Tomography (CT), as well as ultrasound were extensively employed that facilitate the invasive higher-quality and measurable evaluation of cardiac anatomy functions and provide assistance to diagnose, disease monitor, treatment plan and prognosis. Cardiac image segmentation was considered as a critical stage in several applications, which breaks down the images into numerous portions of semantically significant portions, for which the quantitative measures were extracted like myocardial masses, wall thickness, Left Ventricle (LV) volume, Right Ventricle (RV) volume, ejection fraction (EF), etc.

Semantic segmentation represents the classification of all the pixels that belong to specific labels. The clustering process divides the images collectively those were of identical object classes. Since all the pixels of images were classified based on a specific category, this belongs to the form of pixel-level prediction. This semantic segmentation receives the input as images and provides the output as classified regions and structures (inclusive of line segments, curve segments, circles, etc.). Most of the time, the image processing was done is with filters, gradient information, color information, etc. It provides an understanding of the image at the pixel level. The differentiation of objects in the traditional methods is generally a hectic process.

A recognized structure towards medical image segmentation is the U-Net that was formulated by Olaf Ronneberger [10] by employing the idea of deconvolution. This architecture was constructed in a well-designed manner of FCNN (Fully Convolutional Neural Network). U-Net possesses specific benefits by possessing architecture with skip connections amid dissimilar stages in the network [13],[14]. Similarly, U-Net uses few alterations for overcoming the trade-offs amid localizations as well as employment of contexts. This variation increases because a larger-sized patch requires extra pooling layers, moreover it shall decrease the localization accuracy. Alternatively, a small-size patch shall only perceive smaller input contexts.

Besides, a dissimilar feature of architectural design was formulated, i.e., attention. Attention conveys the target location correctly and also rises the region of interest (ROI). The main objective was in increasing the representation power with the help of the attention process, i.e., concentrating on the significant feature as well as by suppressing needless features. This work employs a network module [26] that was commonly termed as the Convolutional Block Attention Module (CBAM). As the convolution operations remove the necessary features by the combination of cross-channel as well as spatial information, this proposed unit was used to extract the necessary features over the two principal components, i.e., channel axis and spatial axis. For achieving this, the channel attention module and spatial attention module were successively applied, thereby all the branches could understand what/where to attend in channel axis and spatial axis. Accordingly, this proposed unit effectively assists in data flow inside the networks by learning the information which needs to be emphasized or suppressed.

Combining the cutting-edge improvements over the attention module, this paper proposes a dynamic dual-attention cardiac image segmentation structure based on U-Net, spatial attention module as well as channel attention module referred to as DPA-UNet. The extracted data from the encoded path was employed to gate the skip connection, discarding inappropriate and noisy responses. Spatial attention modules improve the performances of deep networks in picking spatial correlations amid the features. Channel attention modules improve the network representation in picking the explicit relations amid convolutional channels through a contextual gating mechanism.

Background Study

Segmentation of medical images represents the identification of the organ's pixels from the background images like CT images, MRI images, etc., which was found to be the biggest challenge for analyzing medical images that were supposed to convey serious data regarding the shape and volume of human organs. Numerous research works were implemented with automatic segmentation approaches by using existing techniques. Existing structures were constructed over conventional approaches like edge detection filters and arithmetic models.

Deep learning-based image segmentation

This method was observed as an extensively used robust tool for segmenting the medical images, which was extensively employed for separating identical regions as principal components for diagnosing and treating the diseases. Mohammad Hesam Hesamian et al., [1] carried out a serious assessment over prevalent approaches which used deep-learning (DL) methods to segment the medical images. Furthermore, the required challenges were summarized that experienced probable explanations. Deep neural networks (DNNs) were found as one of the effective methodologies in predicting task-specific features [2].

Similarly, CNN was seen to be effectively employed for resolving image segmentation issues [3]-[5], thereby almost all existing research focused only on non-medical tasks and they include frameworks that were not well-suitable for medical imaging or segmenting brain tumors, in specific. In the past two years, there were growing usages of DL in medical domains, and more precisely for segmenting the medical images. This rise was found in the contemporary Brain tumor segmentation challenges (BRATS) that were held in a combination of

Medical image computing and computer-assisted intervention (MICCAI). Mohammad Havaei [6] analyzed several methods based on deep neural network (DNN) architecture applied in segmenting brain images.

Chen et al., [7] provided a detailed assessment of more than 100 cardiac image segmentation procedures that employed DL strategy that encompasses conventional imaging strategies like MRI, CT, ultrasound, etc., including chief anatomic regions of interest like ventricles and vessels. Additionally, the assessments of freely-available cardiac image datasets were encountered for providing a foundation to encourage recent researchers. Furthermore, the challenges and restrictions in the recent DL-based approaches were analyzed including label scarcity, the generalizability of models in dissimilar domains, interpretability, etc., and also discussed possible future research guidelines.

Shervin Minaee et al., [8] provided an elaborative analysis on prevailing research works encompassing wider spectrums of original research over semantic segmentation as well as instance-level segmentation, comprising of fully-convolutional pixel-labeled networks, encoding-decoding methods, pyramidal strategies, recurrent architectures, visual attention frameworks and propagative methods in combative environments. Moreover, the relevant similarities, benefits as well as constraints in DL approaches, were also examined. Additionally, the extensively employed datasets, performances, future research, etc were also elaborated.

Cardiac MRI Segmentation

Accurate segmentation of cardiac images is a significant stage in assessing the functions of the heart. Christian Baumgartner [17] proposed a fully-automatic architecture to segment the ventricular cavities for short-axis MR images of the heart. Dissimilar 2D and 3D CNN frameworks were analyzed accordingly. The experimental process was carried out on ACDC 2017 challenge training datasets that included 100 patients' cardiac MR images, in which the manual segmentation processes were carried out in end frames of diastolic and systolic phases. Moreover, the mean Dice coefficients of 0.950 (LV), 0.893 (RV), and 0.899 (Myocardium) were reported correspondingly possesses an average 1.1 seconds assessment time over the present GPU.

Fabian Isensee [18] formulated a methodology that addressed the prevailing restrictions by the integration of segmentation and medical image classification to fully-automated processing pipelines. Here, ensembles of -based frameworks were incorporated for segmenting the cardiac features including Left and Right Ventricular cavities (LVC and RVC), as well as left ventricular myocardium (LVM) for all the timing instances in the cardiac cycles. Based on these features, normalized Multilayer perceptrons (MLP) were trained as well as random forest classifiers were used for predicting pathological target classes. Moreover, methods based on ACDC datasets were assessed and was found to attain the Dice score of 0.945, 0.908 and 0.905 of LVC, RVC and LVM over cross-validation in the training set of 100 cases, and 0.950, 0.923 and 0.911pertaining to LVC, RVC and LVM over test sets of 50 cases and conveyed 94% classification accuracy in training set cross-validation with 92% on the test set.

MahendraKhened [20] formulated a fully-automated methodology to segment the LV, RV and Myocardium (Myo) in cardiac MR images with the help of tightly connected FCNN. Dense Convolutional Neural Network (DCNN) enables multiple gradients' path flow amid layers in the training phase using feature propagation as well as back-propagation. Using the test set that was 10% of the training set, this image segmentation methodology attained a mean dice score of 0.92, 0.87, 0.86 concerning LV, RV, Myo. Considering automatic diagnosing of cardiac diseases, classification using random forest showed 90% accuracy.

Jay Patravali [21] proposed 2D/3D segmentation pipelines in fully-automatic cardiac MR image segmentation with the help of DCNN. Such methods were trained with the help of ACD Challenge 2017 datasets including 100 prior analyses, and all these contained cardiac MRI images in the ED phase and ES phase. Ultimately, both segmentation methods achieved performances closer to the well-evaluated performance scores concerning distance metrics as well as revealed substantial accuracy corresponding to the required clinical parameters. Furthermore, an elaborative assessment was carried out by incorporating an original dice loss function as well as combinations of several cross-entropy losses.

Marc-Michel Rohé et al., [22] formulated an automated segmentation architecture that depends upon faster recognition procedure that was trained using CNN. Here, this methodology showed the increased speed that shall permit users with a higher quantity of templates in multi-atlas segmentation but with a computationally-tractable manner. This method was assessed over 100 ED and ES MRI image datasets from the STACOM ACDC 2017 challenge.

JelmerWolterink [23] formulated a fully-automated methodology to segment and it classifies the diseases with the help of cardiac cine MR images. A CNN was formulated for instantaneously segmenting LV, RV, and Myo in ED images as well as ES images. The resultant segmented features were employed over Random forest classifier for labelling the patients suffering from dilated and hypertrophic cardiomyopathies, heart failures resulting from myocardial infarctions, right ventricular abnormalities, or the patients who do not have cardiac diseases. This methodology was formulated and assessed with the help of stable datasets encompassing 100 patients' images that were collected from the MICCAI ACDC 2017 challenge.

Xin Yang [24] formulated a universal and fully-automated response for simultaneously segmenting three significant ventricle features. Based on DL-approach, this methodology was formulated using a 3D fully convolutional network (3D FCNN). Further, 3D FCNN was modified using two well-defined sections like (a) transfer learning that was utilized amid the pre-trained C3D models for attaining excellent initializing and suppressing overfitting, (b) enhancing the gradient flows in networks for being advantageous in increasing the performances of the segmentation process, numerous supplementary loss functions were integrated for exposing the initial layers for yielding improved supervision.

Gomathi and Dr.V.Subha [25] aimed in comparing three DL subsets including the UNet, Segnet and FCNN which were being implemented to segment the ventricle regions including LV, RV and Myo in systolic and diastolic phases of the cardiac cycle. The performance analysis was done based on the values obtained using Hausdorff distance as well as Dice coefficients. Among these three methods, UNet architecture provided the best outcome.

Attention Modules

Trinh Le Ba Khanh [27] introduced an enhanced spatial-channel attention gate that reports the restrictions over plain skip connections and could be combined with encoding-decoding networks for efficiently improving the performances of segmentation tasks. Moreover, the evaluation outcomes revealed the spatial-channel attention gate methodology in largely enhancing the segmentation capabilities of U-Net architectures using nominal computation overheads. Results of this evaluation showed the methodology to outperform those existing architectures possessing a dice score of 71.72%.

Long Chen [28] proposed a modified CNN referred to as SCA-CNN which incorporated spatial-channel wise attention over CNN. For captioning the images, SCA-CNN automatically alters the sentence generation contexts over multiple-layered feature maps, as well as encoding, thereby locating the location of these visual attentions. The evaluation of this formulated SCA-CNN methodology was done over three state-of-the-art datasets like Flickr8K, Flickr30K as well as MSCOCO.

Peng Zhao [30] proposed a DL-based semantic image segmentation network that was termed as SCAU-Net formulated based on recent prevailing research works in medical images. SCAU-Nets possessed an encoding-decoding-styled symmetric arrangement that was combined using spatial-channel attentions as a removable module. The foremost objective here was in enhancing the local features and restraining inappropriate features in spatial-channel attentions. The assessments and evaluations carried out using gland datasets GlaS and CRAG revealed this formulated SCAU-Net approach to be better when compared with the conventional U-Net models for segmenting images. Furthermore, this formulated methodology revealed a 1% increase in Dice score together with a 1.5% increase in Jaccard score.

YutongCai [31] made efforts for eliminating the semantic ambiguities of skip connections with the addition of attention gate (AG), also employed attention methodologies for combining local features with their concerned global feature, moreover modelled the dependences amid channels as well as used multi-scale predictive fusions for utilizing global data in dissimilar scales. On assessment with the well-evaluated existing networks, this approach obtained increased performance in segmenting the images and also introduced a smaller number of parameters.

Foremost features of this research

(1) A novel Spatial and Channel attention UNet (DPA-UNet) methodology have been formulated to attain precise segmentation of cardiac images. The convolutional multi-attentive unit has been employed for extracting the image

features as well as for generating the resulting cardiac image segmentation map. This is done with the help of max pooling and average pooling operations during down-sampling.

(2) The proposed unit inserts the spatial as well as channel attention modules with the help of concatenation procedure. It is carried out during the upsampling to additionally improve the feature representation capabilities. Further, the UNet network that was formed on this unit considerably enhances the cardiac MRI segmentation process.

This paper is structured follows. The articulation and features of the proposed methodology were discussed in Section 2. An elaborative discussion on the simulation results was elaborated in Section 3. Section 4 concludes the paper.

Proposed Architecture

The schematic articulation of this proposed methodology could be seen in Fig.1. The subsequent sections discuss the explanations of individual blocks of this proposed approach.

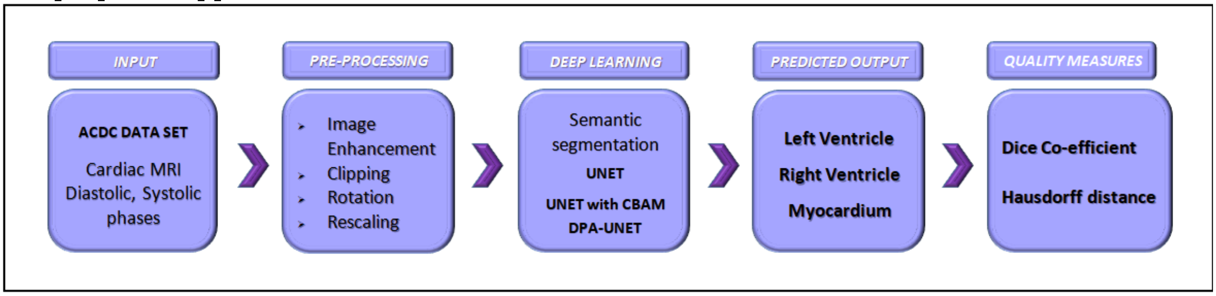

Fig.1 Overall structure of the Proposed Methodology

Dataset

This section provides a brief overview of the dataset. Marc-Michel Rohé [22] proposed an Automatic cardiac diagnosis competition (ACDC) as a novel dataset, which paved a way for the organization of a worldwide MICCAI challenge in 2017. The dataset's vibrancy and strong connections to conventional medical concerns suggest that machine learning approaches might be used to comprehensively interpret cardiac data gotten from MRI. When compared with other prior datasets in cardiac illness, ACDC offered a larger influence due to manual segmentation of RV, LV cavities as well as myocardium (especially epicardial morphology). Individuals with hypertrophic cardiomyopathy (HCM), dilated cardiomyopathy (DCM), altered LV ejection fraction (MINF) in myocardial infarction, anomalous right ventricle, and patients without heart illness are all represented in the ACDC (NOR).

Data Preprocessing

Data preprocessing is the first step in segmentation; it is a standard and common study procedure. Not only does this improve the prediction accuracy, but it also advances the potential for simplifying large amounts of data stability. It is nothing more than an encased process of data augmentation. It is done to increase the dataset's size artificially. It begins with a horizontal and vertical flip. It was then rotated to a degree of ten. The images were rescaled to 1 mm spacing using pixel spacing information. Further, contrast enhancement was increased by a factor of 1.5. The system's endurance was influenced by several factors including scan settings, age, gender types, and the severity of health conditions. Images in the databases are available in a variety of formats and sizes. As a result, they are converted to PNG format with 240x240 dimensions for better display.

Segmentation Architectures

After preprocessing of data, segmenting the regions including RV, LV, and Myo was performed on the phases of systole and diastole using deep learning algorithms. The overview of the algorithms was discussed in subsequent sections.

UNet

UNet was typically constructed with a combination of encoders and decoders. The encoder was employed for retrieving a group of characteristics of the input image, and the decoder was subsequently employed for restoring the group of characteristics to corresponding images. However, it has been found that as the encoder stage progresses, the feature characteristics are lowered and the number of features shrinks. Features were generated later in the decoder step. This network was constructed with concatenated layers, thereby the terminal section of encoders was coupled with lower levels of decoders and the initial section of encoders was coupled to higher levels of functions. This is simple to grasp with a limited training set and its architecture works well. Time elapse of training set for multiple layers and larger images incurred somewhat higher CPU memory utilization that was considered as the main drawbacks of this architecture.

UNet with Convolutional Block Attention Module

CBAM refers to a simpler but efficient attention model concerning feedforward convolutional neural networks. With intermediate maps of feature, this unit successively assumes the attention map over two distinct dimensions, channel/spatial, subsequently, the attention map was multiplied with input feature maps to adaptively refine the features. As CBAM was found to be lightweight as well as a common unit, this unit could be combined with other CNN seamlessly that possess insignificant overhead and it could be trained easily together with the original CNN.
Considering the feature map $F \in R^{C\times H\times W}$ in the data side, CBAM successively assumes both the 1D channel attention maps $M_c \in R^{C\times 1\times 1}$ as well as the 2D spatial attention maps $M_s \in R^{1\times H\times W}$. The entire attention process was elaborated using the equation below,

$$F' = M_c(F) \otimes F, \quad F'' = M_s(F') \otimes F' \qquad (1)$$

Here, $\otimes$ corresponds to the unit-wise multiplication. In this multiplication stage, the attention value was copied completely. Moreover, channel attention value was copied over spatial dimensions, and so on. Likewise, F” corresponds to the finalized refined output.

Initially, aggregated spatial information in feature maps was assessed with the help of average and max pooling operations, producing dissimilar spatial context descriptors, F^c_{avg} as well as F^c_{max} that correspond to both the features individually. These two descriptors were subsequently directed towards shared networks for producing the needed channel attention maps $M_c \in R^{C\times1\times1}$. These shared networks were included with a multilayer perceptron possessing a single hidden layer. For reducing parameter overheads, hidden activation sizes were fixed to $R^{C/r\times1\times1}$, and here r represents the reduction ratio. The shared networks being applied in all descriptors, the vectors of output features were integrated by element-wise summations.

Spatial attention maps with the help of inter-spatial feature relationships were created. The spatial attention mainly focuses on the exact location of the informative part that was the complement for channel attention. For computing spatial attention, average-pooling as well as max-pooling, were initially applied at the channel axis and they were concatenated for generating effectual feature descriptors. To highlight the informative region, pooling processes could be applied over the channel axis and this was found to be largely efficient. Over concatenated feature descriptors, a single convolution layer was applied for generating the spatial attention maps $M_s(F) \in R^{H\times W}$ that typically encoded the region that was to be emphasized or suppressed. Aggregate channel information of feature maps using pooling operations generates dual 2D maps including $F_s avg \in R^{1\times H\times W}$ as well as $F_s\ max \in R^{1\times H\times W}$. All of them signify both the features in the channel and were subsequently concatenated and convolved using the convolution layer, thereby generating needed 2D spatial attention maps.

DPA-UNet

UNet is considered to be significant in segmenting medical images owing to the multi-scale skip connection as well as the trainable up-convolution layer. However, one of the specific features is the inclusion of skip connection which concatenates encoding features concerning the concerned decoding feature that makes the successive process to be effective. On the integration of location data of encoding paths as well as contextual data of decoding paths, the overall data attained is essential for achieving better segmentation maps, thereby improving network performances

Network Architecture.

The process first starts by inputting an image into the architecture. Then, convolution takes place with the max-pooling process in the channel attention gate while average pooling takes place in the spatial attention gate in alternative and vice-versa, but it is not used altogether by each gate. Each convolution takes an input of the precedent pooling process. The spatial attention unit increases the deep learning networks’ performances on picking the spatial correlations amid the features. Channel attention increases the network’s representations in picking the relationship amid the convolutional channel using a contextual gating mechanism. This mechanism takes place in the downsampling phase where the features are gathered, again it is concatenated along with the features gathered during the upsampling process. As a result, every feature gets extracted from the architecture. These features help us to analyze the region of disease in the heart without any missing information. The working mechanism of the DPA-UNet is represented in Fig.2.

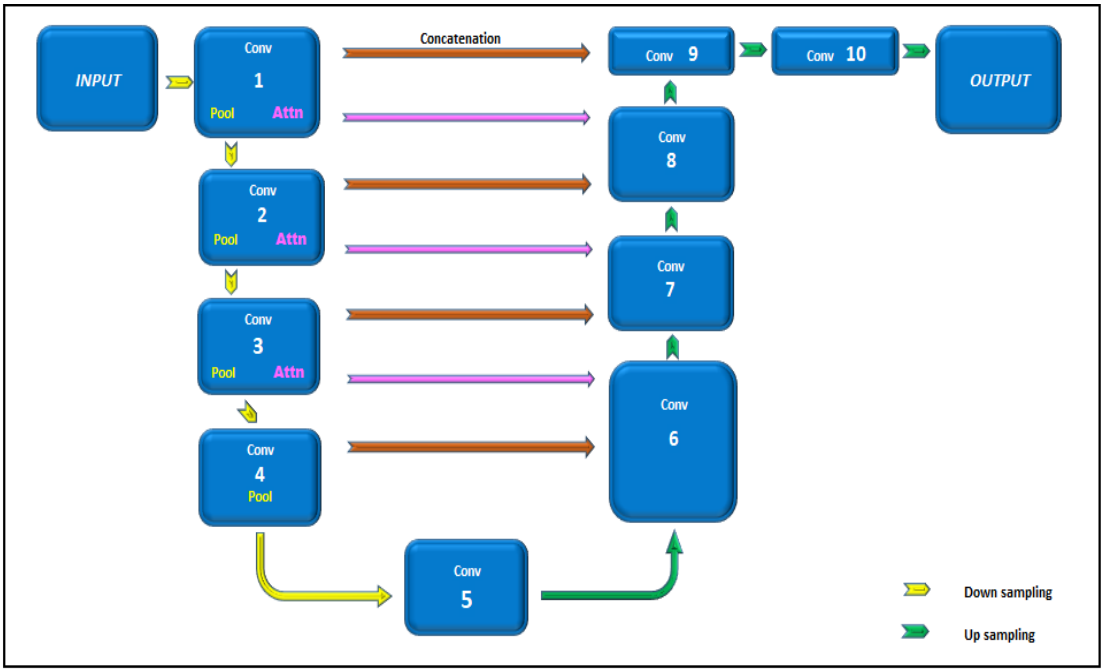

Fig.2. Working mechanism of DPA-UNet

Detail Preserving Attention Block

More precisely, this proposed method integrated the attention gates possessing two components over the skip connection in encoder features and the orresponding decoder features in the UNet (Fig.2). First, spatial Attention Gate (sAG) takes place which is continued by the channel Attention Gate (cAG). Using scAG over conventional encoding-decoding UNet provides us with the intermediate feature maps of the downsampling paths. This is largely applicable in solving image segmentation issues. Furthermore, sAG dynamically focuses on the region of interest, but cAG dynamically understands the representational meaning of regions.

This proposed model includes an attention technique for highlighting the useful information in channel dimensions as well as spatial dimensions of features that will be beneficial in segmenting the images. These attention modules are banded alternately using two pooling techniques, the average and the max pooling. The average pooling process brings out the common features, while the max-pooling process brings out the dominant features. The common features are the basic characteristics of the data segments, while the dominant features are the specific characteristics of the data segments. The feature map strokes are used to describe the channel attention features. The features of spatial attention are described using location-based strokes. In this section, the alternate bands of attention modules and pooling layers are incorporated into the equation. When compared with other existing methods, this structure provides perfect segmentation. It eliminates the possibility of feature overfitting as well as yields the highest accuracy with reduced feature losses. The working mechanism of the attention blocks could be understood from Fig.3 (a) as well as Fig.3 (b). In fig.3 (a), channel attention is combined with the average pooling, while spatial attention is combined with maximum pooling. As a result, all the features were properly segmented, possessing the highest accuracy for the segmentation process. In fig.3 (b), the process reverses itself.

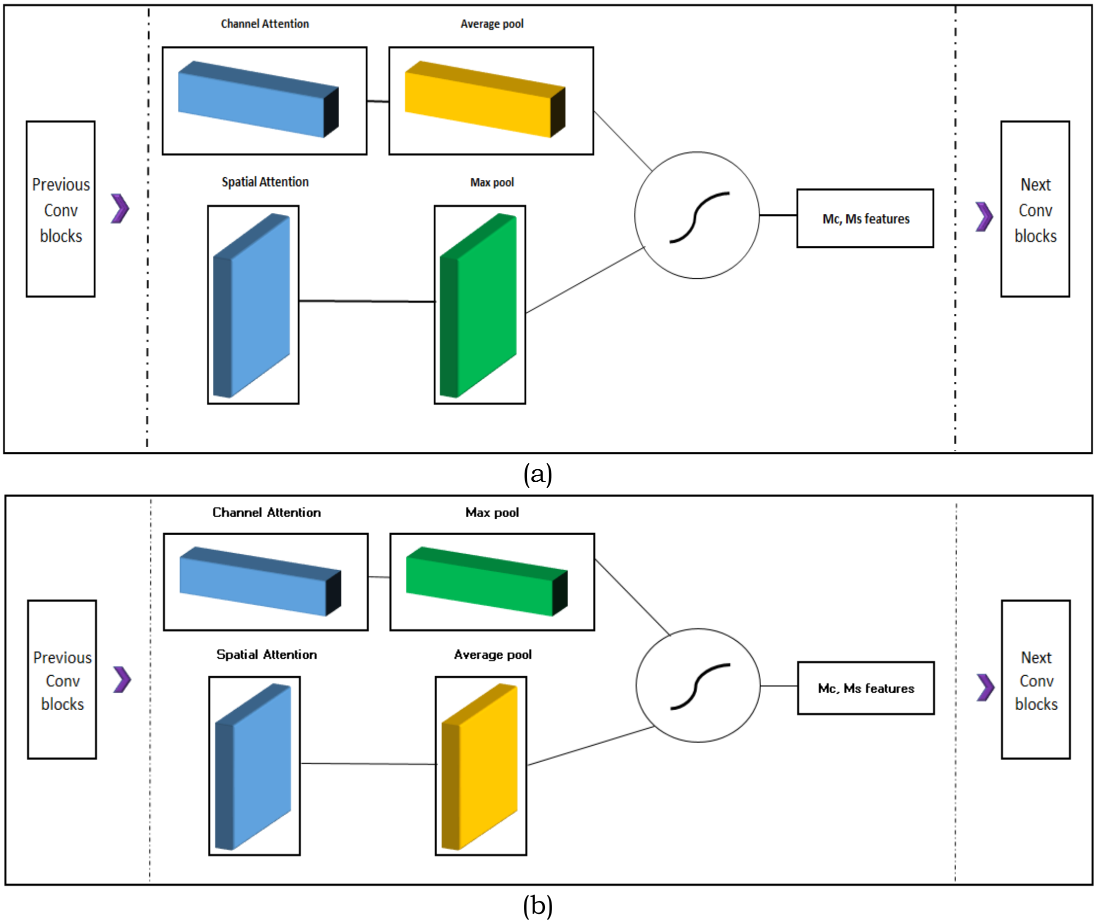

Fig.3. (a) Process flow of attention module1 and attentionmodule 3 (b) Process flow of attention module2

Comparison state of art

UNet method is common in the field of MRI segmentation. Moreover, the time consumption is less compared to the CBAM method. However, there are some flaws in this method. To make the entire process effective, the CBAM method can be used. This is because the attention mechanism is blended with it. The

Proposed DPA-UNet comprises a channel and spatial attention. The researchers channelize pooling under the above-mentioned attention. It results in is no overlapping of features. Thereby, is no image feature is missed.

Pseudocode for the Proposed DPA-UNet Methodology

The pseudocode of the proposed methodology was explained below in Fig.4.

Input: Cardiac MRI Diastole, Systole Images
Output: Left Ventricle, Right Ventricle, Myocardium
Procedure

```
Function ca() //Channel Attention
MLP  = Dense(channel, Relu,)
a=Average Pooling(channel feature)
m=Max Pooling(channel feature)
Return(features)
Function sa() //Spatial Attention
a =  mean(spatial feature)
m = max(spatial feature)
Return (features)
Read the Training images (Diastole, Systole) for 100 patients
Read the mask images
Fit the training data into the DPA-UNet model
```

Downsampling

```
Training Process
C1=conv(32,(3,3),Relu)
P1= maxpool(C1)
Att1=ca(P1)
//Channel Attention :Average Pool, Spatial Attention: MaxPool
C2=conv(64,(3,3), Relu)
P2= maxpool(C2)
Att2=sa(P2)
Channel Attention: Max Pool, Spatial Attention- Average Pool
C3=conv(128,(3,3), Relu)
P3= maxpool(C3)
Att3=ca(P3)
//am – Channel Attention :Average Pool, Spatial Attention: MaxPool
C4=conv(256,(3,3), Relu)
P4= maxpool(C4)
C5=conv(512,(3,3), Relu)
P5= Maxpool(c5)
```

Upsampling

```
U6=concatenate(c5,c4,att3)
C6 = conv(256, (3, 3), Relu)(U6)
U7 = concatenate(c6, c3, att2)
C7= conv(128, (3,3), Relu)(U7)
U8= concatenate(c7, c2, att1)
C8 = conv(64,(3,3), Relu)(U8)
U9 = concatenate (c8,c1)
C9= conv(32,(3,3), Relu)(U9)
C10=conv(4,(1,1), softmax)(C9)

Output predicted with 4 classes(LV,RV,MYO,Background)
```

Fig.4. Pseudocode for DPA-UNet

Experimental Analysis

This section deals with the experimentations carried out in cardiac images for deep learning algorithms. The performances were calculated using the Dice coefficient and Hausdorff distance. DPA-UNet was compared with the existing algorithms which depict that this proposed method outperforms the other two existing strategies.

Dice coefficient (DC)

DC was identified to be identical to Jaccard's index, and also DC double-counts the intersection (TP). Dice score was typically employed for quantifying the performances of methodologies for segmenting the images. DC could be mathematically expressed as,

Dice Coefficient (DC) = $\frac{2TP}{2\,TP+FN+FP}$ (2)

Here, TP signifies true positive, TN represents true negative, FP corresponds to false positive, and FN signifies false negative.

Hausdorff Distance (HD)

HD estimates the distance of the two subsets with respect to each other. Normally, this corresponds to the maximum distance from one point in a set with respect to the nearest point of another set. Let, A and B be the two non-empty subsets in metric spaces possessing distance metrics D, therefore HD could be mathematically expressed as,

$D_{H(A,\,B)} = \max\{h(A, B), h(B, A)\}$ (3)

Here, A represents Set 1 and B signifies Set 2. The systole and diastole datasets were incorporated with 80% training set as well as 20% test set.

Table 1
Quality Measures of Dice Coefficient (DC)

Methods	Dice Coefficient					
	Diastolic Phase			Systolic Phase		
	LV	RV	MYO	LV	RV	MYO
2D UNet	0.964	0.933	0.897	0.917	0.886	0.903
UNet with CBAM	0.968	0.941	0.899	0.93	0.89	0.91
DPA-UNet	0.976	0.954	0.905	0.942	0.9105	0.928

Table 2
Quality Measures of Hausdorff Distance (HD)

Methods	Hausdorff Distance					
	Diastolic Phase			Systolic Phase		
2D UNet	6.7	10.5	8.4	9.5	14.8	10.2
UNet with CBAM	6.5	10.1	8.3	7.6	12.4	11.8
DPA-UNet	6.1	9.8	8.1	6.1	11.01	8.1

Tables 1 & 2 represent the outcomes from DC and HD assessments in all the three regions namely the left ventricle, right ventricle and myocardium that were calculated separately for methods including UNet, CBAM and DPA-UNet diastolic phase and systolic phase respectively. From both the tables, it was inferred that this proposed methodology DPA-UNet outperformed the peers in terms of the segmentation process with effective results.

Figs. 5 & 6 represent the overall outcomes of performance of all the three network assessments in both the phases (systolic and diastolic) respectively. From graphical representations, it could be inferred that DPA-UNet outperformed the other two methods namely UNetand CBAM in every aspect achieving the highest results in Dice coefficient and Hausdorff distance for both the phases (diastolic and systolic).

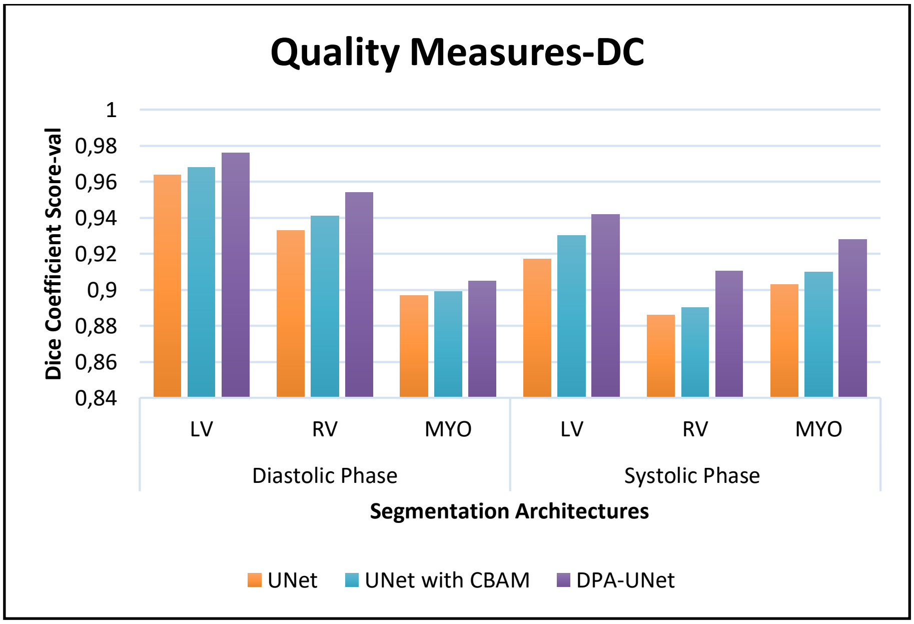

Fig.5. Evaluation Metric–DC

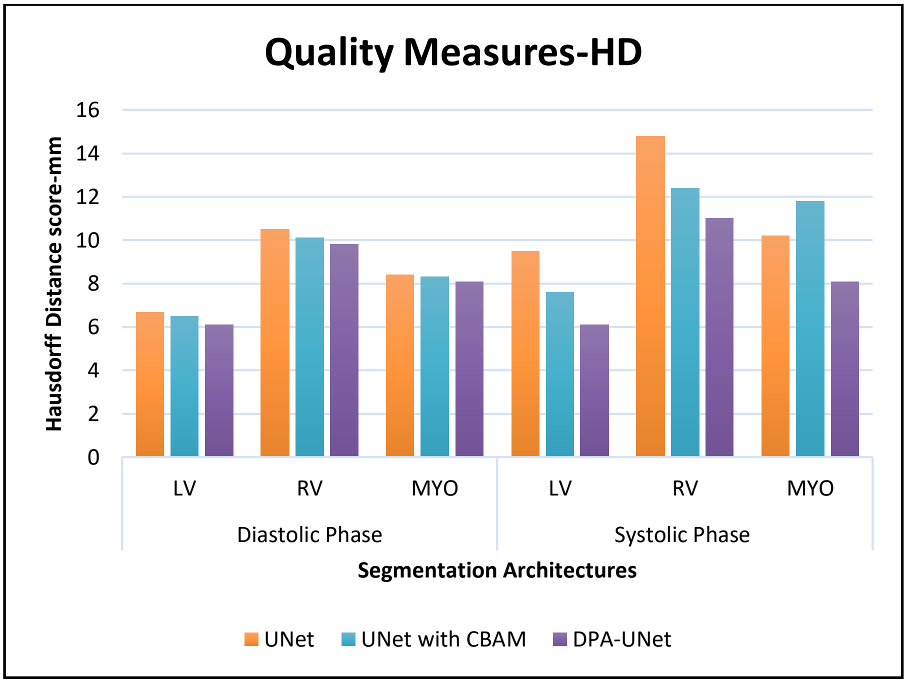

Fig.6. Evaluation Metric–HD

Fig.7 shows the segmentation of all three areas, LV (Red), RV (Blue), and myocardium (Green) individually in both the phases that were accomplished utilizing different architectures discussed above. From fig.5., it could be inferred that DPA-UNet architecture segmented all the three regions separately without any data losses in an effective way.

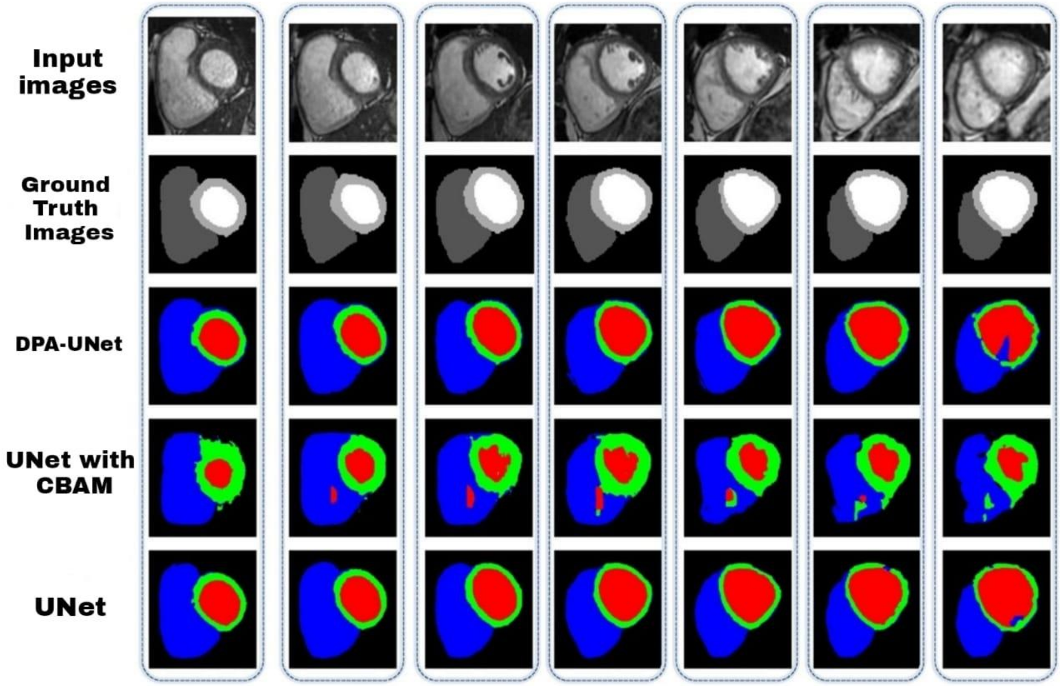

Fig.7. (a) Segmentation of all the three regions in diastolic phase

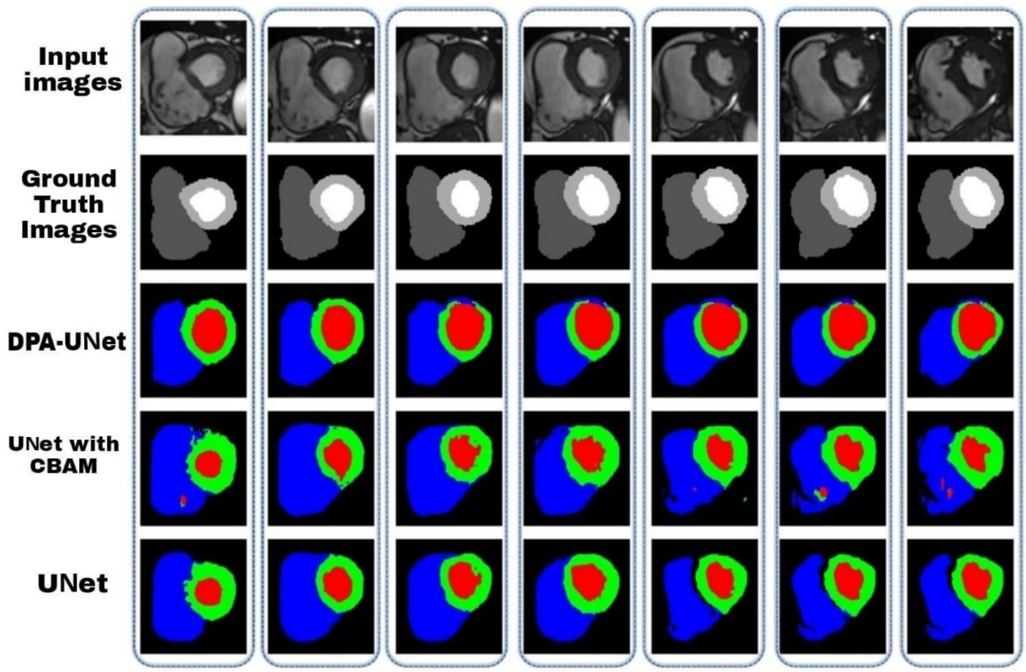

Fig.7. (b) Segmentation of all the three regions in systolic phase

Table 3 and Fig.8 represent the training time in seconds for all the three deep learning methods in both phases. It was inferred that this proposed method DPA-UNet achieved less time with accurate results.The prediction time for all three methods was observed as 0.015 seconds.

Table 3
Training Time for deep learning methods

Training Time (Seconds)		
Methods	Diastolic Phase	Systolic Phase
UNet	299	244
UNet WITH CBAM	394	310
DPA-UNet	312	293

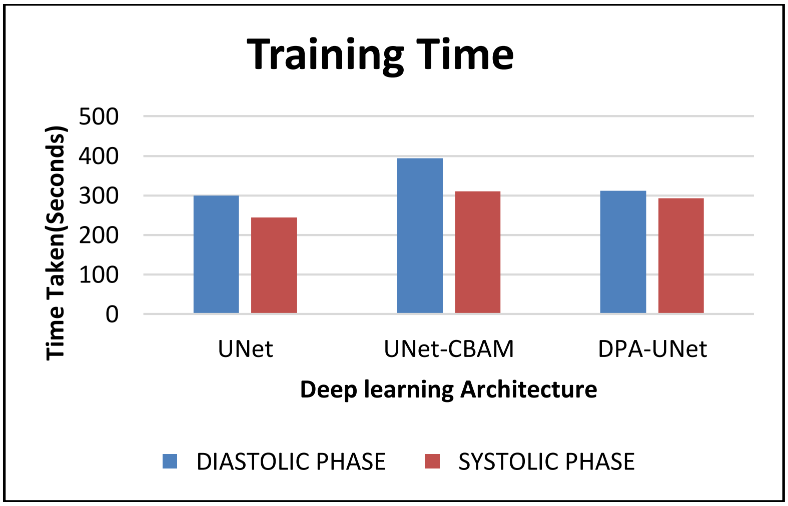

Fig.8. Training time for all the three deep learning methods

Table 4 represents the prediction time of all three deep learning methods. It was inferred that the proposed methodology DPA-UNet outperformed other methods in segmenting the disease regions in cardiac regions in less time with more accurate results.

Conclusion

In this paper, an effective methodology concerning deep learning structure, the DPA-UNet is proposed that particularizes attention modules for extracting the ventricle regions. The summarization of this work is the concatenation of pooling with alternative combinations of the channel and spatial attention that have resulted in competitive results. This method shall provide a new way to understand the representation of deep learning structures. The researchers apply attention-based features with the refinement of two distinct alternate combinations of pooling. Overall, this method achieved competitive performances in the extraction process. The proposed approach has been found and expected to possess greater potential for application over additional modules in medical image segmentation. Simulation results clearly show that this proposed DPA-UNet architecture possesses increased performance for effectively segmenting the ventricle regions in cardiac MRI images.

References

1. Mohammad Hesam Hesamian, WenjingJia, Xiangjian He, Paul Kennedy, “Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges”, Journal of Digital Imaging, https://doi.org/10.1007/s10278-019-00227-x, 2019
2. Bengio, Y., Courville, A., Vincent, P.: Representation learning: A review and new perspectives. Pattern Analysis and Machine Intelligence, IEEE Transactions on 35(8), 1798–1828 (2013)
3. Alvarez, J.M., et al.: Road scene segmentation from a single image. In: Computer Vision–ECCV 2012, pp. 376–389. Springer (2012).
4. Hariharan1, B., et al.: Simultaneous detection and segmentation. arXiv preprint arXiv:1407.1808 (2014).
5. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3431–3440 (2015)
6. Mohammad Havaei,” Deep learning trends for focal brain pathology segmentation in MRI”, Machine Learning for Health Informatics, arXiv:1607.05258v3 [cs.CV] 24 Jan 2017.
7. Chen Chen1, Chen Qin, HuaqiQiu, Giacomo Tarroni, JinmingDuan, Wenjia Bai and Daniel Rueckert, “Deep Learning for Cardiac Image Segmentation: A Review” Frontiers in cardiovascular medicine, Volume 7, article 25, March 2020.
8. Shervin Minaee, Yuri Boykov, FatihPorikli, Antonio Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos, “Image Segmentation Using Deep Learning: A Survey”, arXiv:2001.05566v5 [cs.CV] 15 Nov 2020.
9. Shaoqiong Huang, “Medical image segmentation using deep learning with feature enhancement”, ET Image Process., 2020, Vol. 14 Iss. 14, pp. 3324-3332.
10. Olaf Ronneberger, “U-Net: Convolutional Networks for Biomedical Image Segmentation”, Springer International Publishing Switzerland 2015, MICCAI 2015, Part III, LNCS 9351, pp. 234–241, DOI: 10.1007/978-3-319-24574-4_28.
11. Dominik Müller, “MIScnn: a framework for medical image segmentation with convolutional neural networks and deep learning”, Müller and Kramer BMC Med Imaging (2021) 21:12
12. NAHIAN SIDDIQUE, “U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications”, IEEE, VOLUME 9, 2021.
13. 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.

14. Jonathan Long, "Fully Convolutional Networks for Semantic Segmentation", IEEE, 978-1-4673-6964-0/15
15. Liset Vázquez Romaguera, Francisco Perdigón Romero, "Left ventricle segmentation in cardiac MRI images using fully convolutional neural networks", February 2017 DOI: 10.1117/12.2253901.
16. Chen Li, Yusong Tan, "ANU-Net: Attention-based nested U-Net to exploit full resolution features for medical image segmentation", Computers & Graphics, MAY 2020.
17. Christian F. Baumgartner, "An Exploration of 2D and 3D Deep Learning Techniques for Cardiac MR Image Segmentation", LNCS 10663, pp. 111–119, 2018.https://doi.org/10.1007/978-3-319-75541-0_18
18. Fabian Isensee, "Automatic Cardiac Disease Assessment on cine-MRI via Time-Series Segmentation and Domain-Specific Features", LNCS 10663, pp. 120–129, 2018.https://doi.org/10.1007/978-3-319-75541-0_13
19. Yeonggul Jang, "Automatic Segmentation of LV and RV in Cardiac MRI", LNCS 10663, pp. 161–169, 2018.https://doi.org/10.1007/978-3-319-75541-0_17
20. MahendraKhened, "Densely Connected Fully Convolutional Network for Short-Axis Cardiac Cine MR Image Segmentation and Heart Diagnosis Using Random Forest", LNCS 10663, pp. 140–151, 2018.https://doi.org/10.1007/978-3-319-75541-0_18
21. Jay Patravali, "2D-3D Fully Convolutional Neural Networks for Cardiac MR Segmentation", 7, LNCS 10663, pp. 130–139, 2018.https://doi.org/10.1007/978-3-319-75541-0_14
22. Marc-Michel Rohé, MaximeSermesant, Xavier Pennec. Automatic Multi-Atlas Segmentation of Myocardium with SVF-Net. STACOM: Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges, Sep 2017, Québec, Canada. pp.170-177, ff10.1007/978-3-319-75541- 0_18ff. ffhal-01575297f
23. Jelmer M. Wolterink, "Automatic Segmentation and Disease Classification Using Cardiac Cine MR Images", LNCS 10663, pp. 101–110, 2018.https://doi.org/10.1007/978-3-319-75541-0_11
24. Xin Yang," Class-Balanced Deep Neural Network for Automatic Ventricular Structure Segmentation", LNCS 10663, pp. 152–160, 2018.https://doi.org/10.1007/978-3-319-75541-0_16.
25. G. Gomathi, Dr.V.Subha, "Semantic Segmentation of Ventricular and Myocardium Regions in Cardiac MRI", ISSN: 0011-9342, Issue: 9, Pages: 8510-8522, 2021.
26. Sanghyun Woo," CBAM: Convolutional Block Attention Module", arXiv:1807.06521v2 [cs.CV] 18 Jul 2018.
27. Trinh Le Ba Khanh, "Enhancing U-Net with Spatial-Channel Attention Gate for Abnormal Tissue Segmentation in Medical Imaging", Appl. Sci. 2020, 10, 5729; doi:10.3390/app10175729reeLong Chen, "SCA-CNN: Spatial and Channel-wise Attention in Convolutional Networks for Image Captioning", IEEE
28. Jingfei Hu, "SA-Net: A scale-attention network for medical image segmentation", PLOS ONE | https://doi.org/10.1371/journal.pone.0247388 April 14, 202

29. Peng Zhao, "SCAU-Net: Spatial-Channel AttentionU-Net for Gland Segmentation", Frontiers, published: 03 July 2020, DOI: 10.3389/fbioe.2020.00670.
30. YutongCai, " MA-UNet: An improved version of UNet based on multi-scale and attention mechanism for medical image segmentation",arXiv:2012.10952v1 [eess.IV] 20 Dec 2020