Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 6 (2023)

An Investigation of Deep Supervised Approaches for Ventricle Region Segmentation using Cardiac MRI

G. Gomathi [1], Dr. V. Subha[2] , Dr. A. Manivanna Boopathi[3]

Research Scholar [1], Assistant Professor[2],Professor[3].

Department of Computer Science and Engineering, Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu, India[1,2]

Department of Electronics and Instrumentation Engineering,Saveetha Engineering College, Chennai, Tamilnadu, India[3]

Abstract

The precise delineation of ventricle regions through Cardiac MRI segmentation is a vital aspect of quantitative analysis of cardiac function. This facilitates the early detection and follow-up of many heart conditions. The goal of this investigation is to evaluate the most advanced deep learning methods presently in use, such as UNet, Segnet, FCN(Fully Convolutional Network), UNet++, and UNet-CBAM. The techniques are capable of extracting significant elements from the input images without receiving any manually constructed characteristics. The integration of deep learning techniques has brought notable improvements to the field of cardiac MRI segmentation. According to the evaluation metrics, all five models were highly meticulous at segmenting ventricle sections, whereas UNet-CBAM outperformed than other four models. Numerous studies have demonstrated that deep learning-based segmentation techniques perform more accurately and effectively than conventional segmentation techniques. Overall, the adaptation of deep learning models exposed a raise in the semantic segmentation framework of Cardiac MRI.

Keywords:*UNet, SegNet, FCN, UNet++, UNet-CBAM, deep learning*

1. Introduction

Different heart disorders are particularly important in the present day since they have a high death rate. To stop and manage Cardio Vascular Disease (CVD), early identification is crucial [1-3]. The World Health Organisation estimates that 16% of all deaths globally were caused by cardiovascular diseases (CVDs). There were 8.9 million more fatalities from this illness in 2019 than there were in 2018. By 2035, there will be 30% more persons with CVDs, according to forecasts [4]. To diagnose diseases and calculate risk, the heart function must be examined using MRI scans [5]. Heartbeat images can be seen in detail using cardiac magnetic resonance imaging. This can aid the physician's research into the composition and operation of the heart. The process of dividing an image into semantically (i.e., anatomically) significant parts, such as left and right ventricles are known as segmentation in the context of cardiac MRI.

With the use of deep learning (DL) techniques, the field of medical image processing has seen significant breakthroughs since 2015. DL made outstanding progress, completely altering how medical images are processed and analysed. Because of DL's high generalisation capacity, high performance, and versatility, the scientific community is inspired by this now. The advancement of powerful computers and an abundance of medical data are the sources of inspiration for DL [7–10]. Poor contrast in the myocardium and its surroundings, as well as brightness heterogeneities in the ventricular chambers, are just two of the many challenges with CMR

segmentation that are highlighted in [11]. iii) a constrained CMR resolution, etc. Latterly, CNN-based algorithms achieved exceptional results in medical image segmentation, yet they are unable to meet the challenging standards of segmentation accuracy required by medical applications. In medical image analysis, accurate image segmentation remains a difficult job [12]. Utilisingatrous convolutional layers [13,14], self-attention methods [16,17] and image pyramids [15], some researchers have attempted to solve this issue.

This research is employed to note the similarities and dissimilarities between the performance of various deep learning techniques like UNet, SegNet, FCN and UNet++. These are the techniques mainly used in the field of cardiac image segmentation. This study employs a computational analysis approach to assess the segmentation accuracy, computational efficiency, and generalization ability of the five models. Due to several factors, including imaging artefacts, intricate anatomical features, and differences in picture quality, accurately segmenting the heart ventricle region from MRI scans is difficult. As a result, this work offers insightful information about the utilization of deep learning approaches for cardiac ventricle segmentation and emphasises the significance of choosing the right model depending on the intended result. The results of this study can help with the development of precise and effective segmentation systems, which will improve cardiovascular disease diagnosis and treatment planning.

2. Literature Review

The complex anatomical components and the wide range in picture quality make it difficult to segment the heart ventricle region from medical images. Deep learning approaches have recently demonstrated considerable promise for problems like cardiac ventricle segmentation in the area of classifying medical images.Due to UNet architecture's exceptional performance in maintaining spatial information and lowering the number of parameters, it has been widely employed in medical picture segmentation applications since it was first proposed by Ranneberger et al. in 2015 [20]. Similar to this, the SegNet architecture suggested by Badrinarayanan et al. in 2017 [19] has also demonstrated encouraging outcomes in medical picture segmentation tasks, including cardiac ventricle segmentation.

The Fully Convolutional Network (FCN), developed by Long et al. in 2015 [18], is another well-liked deep learning architecture for image segmentation tasks. FCN is a good choice for medical image segmentation problems since it can do pixel-level segmentation and accept input images of any size. The UNet++ architecture, which Zhou et al.,[21] suggested in 2018, improves on the original UNet architecture's performance by adding several skip routes to the encoding and decoding phases. This change raises segmentation accuracy and enhances feature representation.

Convolutional Block Attention Module (CBAM), a new architecture, was recently unveiled by Woo et al. in 2018 [22]. The channel attention and spatial attention modules in CBAM enable the network to focus on crucial features while suppressing unnecessary ones. Investigation of Deep learning methods has been done as it relates to cardiac ventricle segmentation problems. For instance, Wang et al., [24] assessed UNet and FCN's performance in segmenting the left ventricle from cardiac MRI data and discovered that both models obtained good segmentation accuracy. Similarly to this, Zhang et al. achieved good segmentation accuracy by using the SegNet architecture to separate the left ventricle from cardiac MRI images.

3. Methodology

The general methodology of cardiac image segmentation has three main categories like image pre-processing, image segmentation using deep learning techniques and evaluation of the segmented images using performance metrics.It is illustrated inFigure1.The input image first undergoes preprocessing. The input images from different databases are in different formats and sizes. For better demonstration, they have been converted to PNG format with 240x240 dimensions. Next, the pre-processed images are subject to cardiac segmentation by deep learning techniques. In this study, UNet, SegNet, FCN and UNet++ are employed. After segmentation by these techniques, the resultant image is subjected to evaluation and find out the best technique is prompt for cardiac segmentation. The deep learning techniques employed in this study are elaborated in the following sections.

Figure 1. Process Flow of Semantic Segmentation

A. Cardiac MRI Segmentation using FCN

FCN architecture is considered the best DL (Deep Learning) network for segmentations [18]. It is essentially formed with a series of encoders and decoders, and the complete architecture is depicted in Fig.2.This encoder structure is formed using several layers that are applied after each other with convolution and non-linear activation procedures. In the segmentation task, the encoder conducts encoding of significant image features in common. After that, the complete features attained from bottleneck layers have to be oversampled towards the ground truth image, with decoders performing image segmentation per pixel. Finally, the softmax classifier receives the decoder output and generates the final output. In a fully convolution neural network, FCN32 architecture is used and the image size is reduced and the kernel size is increased in the final convolution of sixth and seventh blocks. At the final stage, the transpose operation has to be performed and then the segmented output is predicted when the Kernel size is reduced to 16x16. As long as every connection are local, FCN is capable of handling an extensive variety of image sizes. An upsampling path is used to enable localization, whereas a downsampling path is used to extract and interpret the context. Transposed convolution is used for upsampling, which is an operation that works in the opposite direction of a convolution and allows the activations to be translated into something meaningfulwith the image size by scaling up the activation size to the same image size.

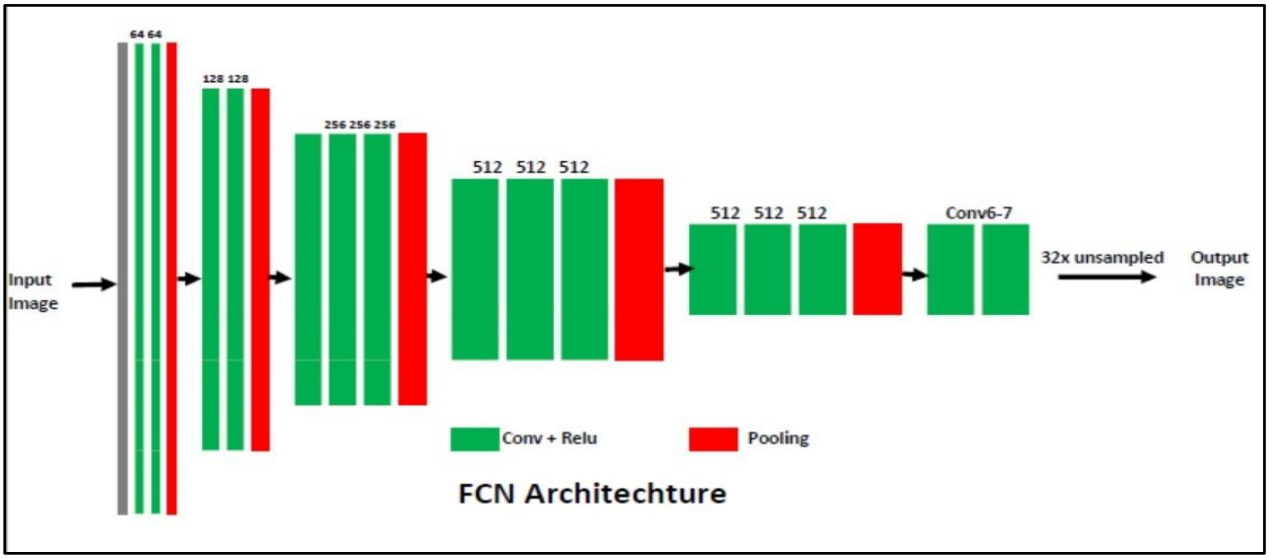


Figure 2. Detailed architecture of FCN[23]

The major factor in choosing CNN is the weight-sharing aspect of this design, which reduces the amount of trainable network variables and enables the network to increase generalisation and prevent overfitting. The fully connected layer being incredibly computationally expensive is the drawback of this system.

B. Cardiac MRI Segmentation using SegNet

SegNet is a deep fully convolutional neural network architecture which is mainly used for semantic pixel-by-pixel segmentation [19]. The encoder, related decoder, and pixel-wise categorization layer make up the

fundamental components of the trainable segmentation engine. Each encoder in the encoder network runs a convolution operation with a filter bank to produce a set of feature maps. The position of the feature map in sub-sampling (grouping indices) was recorded in encoding over the SegNet architecture, so the decoder can construct the sparse feature map while sampling the input with the help of pooling indices. The sparse feature map was subsequently perverted using training filters for yielding the feature map that was further given to the softmax classifier for providing the pixel-by-pixel image segmentation. And also, comprises a deep encoder network. and a layer of decoders-each for the encoders and one pixel-wise classification layer amid them. Max-pooling indices calculated at pooling steps from relevant encoders were delivered to suitable decoders, which conduct deconvolution with a nonlinear up-sampling of the feature map which is obtained as an input. The sparse upsampled feature map was further convolved with training filters for producing denser feature maps in decoders. A SegNet-type decoder may be used for replacing deconvolution filters learned at the original T-Net architecture, resulting in a hybrid architecture with fewer network parameters. The architecture for SegNet is displayed in Figure.3.

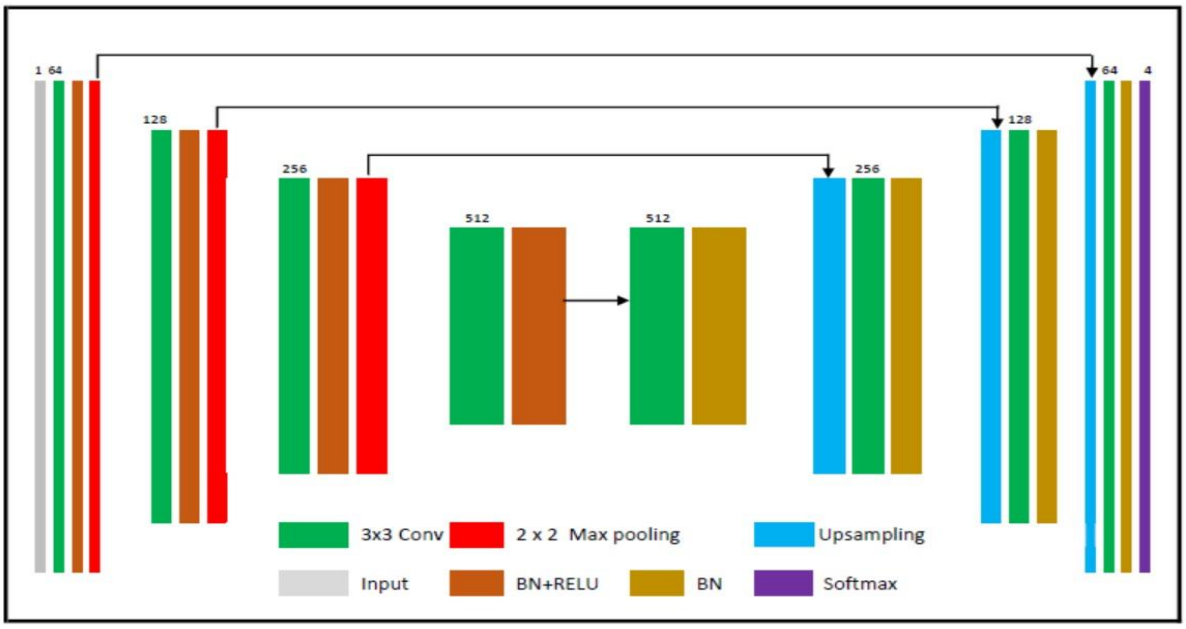


Figure 3. Detailed architecture of SegNet[23]

In SegNet, the image size starts with 240x240 and the filter size varies in every convolution block by 64,128,256,512. This architecture has four blocks like upsampling in the decoder path with the reduced kernel size. When the kernel size reached 64, then the image size gets increased step by step with the multiples of two, and finally, in the softmax layer, the segmented results of cardiac ventricle regions were predicted. The main advantage of this architecture is that it was formulated for being effectual memory and computational time in the course of inferences. Moreover, it possesses a substantial quantity of training parameters when compared with existing architectures and was given end-to-end training with the help of stochastic gradient background.

C. Cardiac MRI Segmentation using UNet

UNet was created and used for the first time to process biomedical images in 2015 [20]. It has a “U” shape structure. Figure 4 highlights the UNet's comprehensive architecture**.** A passage that grows to the right and a path that compresses to the left make up the symmetrical architecture. It is used to extract a set of features from the actual image of the ground. The contracting path is cohesive in accordance with the standard convolutional network architecture. Rarely, two 3x3 convolutions (unpadded convolutions) are used after a rectified linear unit (ReLU) and a 2x2 max pooling operation using stride 2.The total number of feature channels is divided into four for each downsampling step. The down-sampling procedure may work with the image by the following parameters such as 240x240 image size, Filter size=32, Kernel size=3x3, Pooling=2x2, Image size=120x120, Filter size=64, Kernel size=3x3, Pooling=2x2, Image size=60x60, Filter size=128, Kernel

size=3x3, Pooling=2x2, Image size=30x30, Filter size=256, Kernel size=3x3, Pooling =2x2,Image size=15x15, Filter size=512, Kernel size=3x3, Pooling=2x2 respectively.

Each stage of the expanded path includes an upsampling, which is then used to restore the set of characteristics to the appropriate images. Two 3x3 convolutions are used to combine the feature map with a similarly cropped feature map from the constricted path, usually followed by a ReLU. This 2x2 convolution ("up-convolution") lessens the number of feature channels. Cropping was made necessary because each convolution had fewer border pixels. The final layer uses a 1x1 convolution to transfer each of the 64-component feature vectors to the appropriate number of classes. In addition, U-Net reduces the loss functions in the ground image by linking the layers to both the contracting and expansive path.

This architecture's primary benefit is the fact that network input image size is agnostic because of that there are no fully connected layers in this architecture. As a result, the model's weight is reduced and several classes are scaled easily. With a small training set, this is easy to understand, and the architecture is sound. The limitations of this design includes a fairly high CPU memory utilization for training sets with numerous layers and larger images.

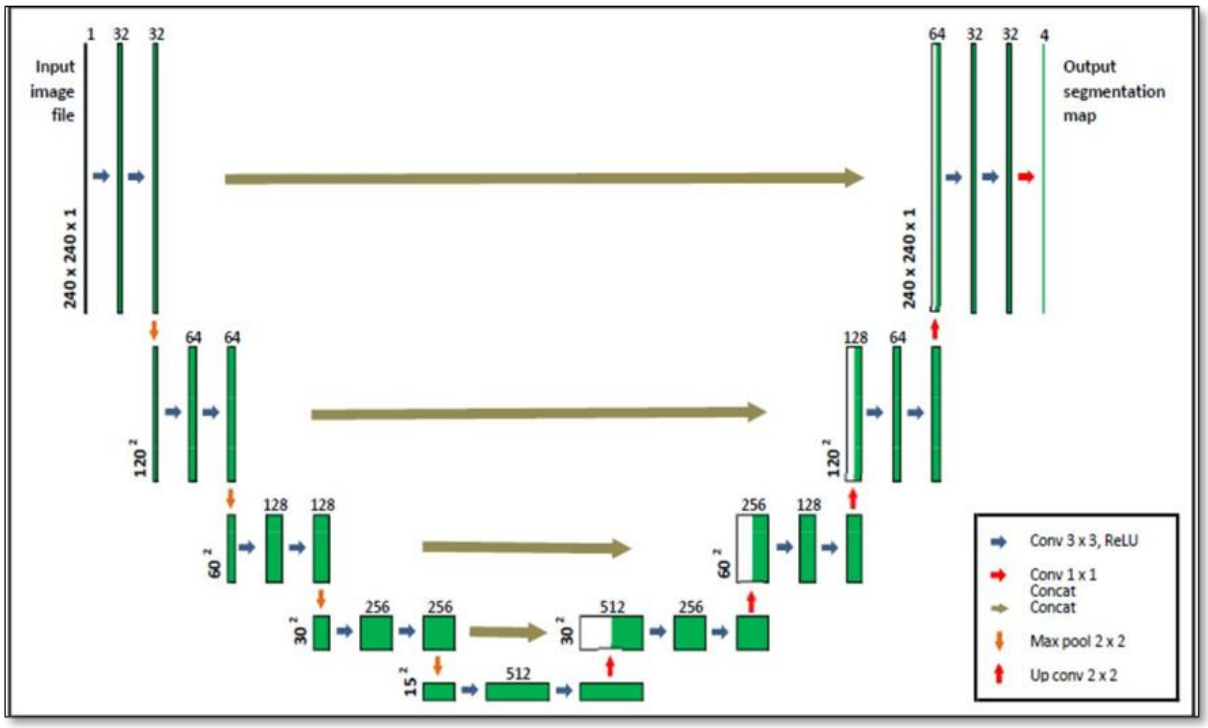


Figure 4.Detailed architecture of UNet[23]

D. Cardiac MRI Segmentation using UNet++

The medical image segmentation architecture UNet++ is more powerful. Through the use of numerous nested, dense skip paths, this design successfully constructs a deeply-supervised encoder-decoder network. A standard overview of this architecture is shown in Figure.5.

The connectivity of the encoder, as well as the decoder sub-networks, are altered through redesigned skip pathways [21]. The decoder in UNet receives the feature maps directly from the encoder without travelling via any sub-architectures, as it does in UNet++, which traverses by a dense convolution block whose size is dependent on the pyramid level.

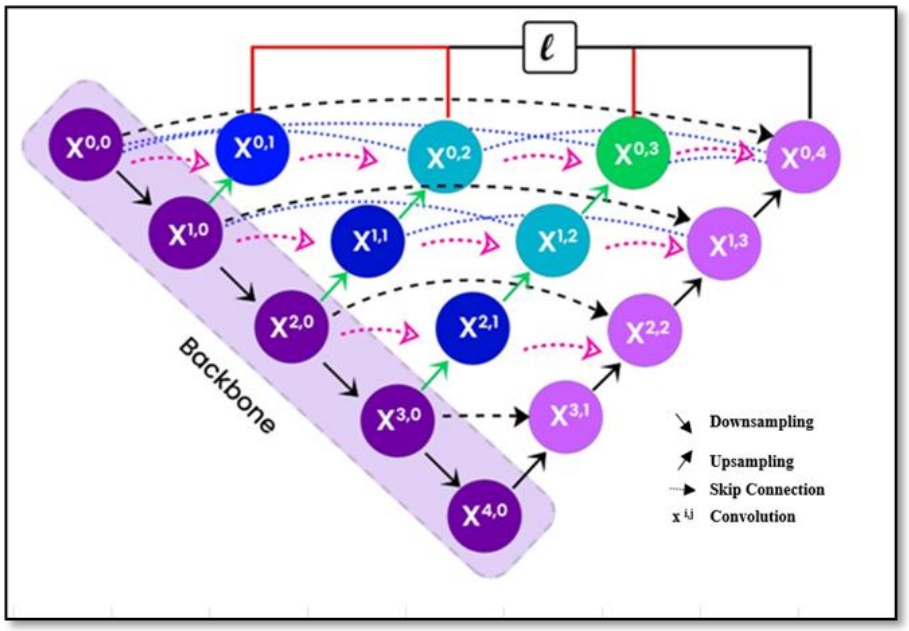


Figure 5. Detailed architecture of UNet++

Using the skip pathway as an example, the concatenation layer, which comes before each convolution layer and mixes up-sampled input from the lower dense block with output from the prior convolution layer of the same dense block, is located between nodes $p^{0,0}$ and $p^{1,3}$. This dense convolution block consists of three convolution layers. The dense convolution block effectively enhances the cognitive relevance of the characteristic maps used by the encoder to that of the features awaited by the decoder. The following is the formalised skip pathway: Let $p^{i,j}$ represent the output, where i is the dense block's convolution layer through the skip pathway as well as j is an encoder's downsampling layer. Calculations are made for the stacked set of feature maps shown as $p^{i,j}$.

$$p^{i,j} = \begin{cases} S(p^{i-1,j}), & j = 0 \\ S\left([p^{i,k}]_{k=0}^{j-1},\ N(p^{i+1,j-1})\right), & j > 0 \end{cases} \quad (1)$$

where [] refers the concatenation layer, N(.) refers an up-sampling layer, and function S(.) refers a convolution process observed by an activation function. Level j = 0 nodes effectively only get one input from the layer above the encoder; level j = 1 nodes effectively get two inputs from the encoder sub-network but at two successive levels; and level j > 1 nodes effectively get j + 1 inputs, where j inputs are the outputs of the previous j nodes in the same skip pathway and the last input is the up-sampled output from the lower skip pathway. Each skip pathway has a dense convolution block, which leads all earlier feature maps to converge together and assemble at the current node.

E. Cardiac MRI Segmentation using UNet-CBAM

Convolutional Block Attention Module (CBAM) CBAM is a lightweight, multi-purpose module that is simple to add to any CNN architecture and that can be trained end-to-end using simple CNNs[22]. Figure.6 exhibits the UNet-CBAM's detailed architecture, and the modules ATTE1, ATTE2, and ATTE3 indicate the channel and spatial attention methods used during the segmentation process.

1. **Attention mechanism:** It is widely acknowledged that human perception depends heavily on attention [23–25]. A significant feature of an individual's visual system is that it never seeks to undertake an entire scene at once. Humans alternatively employ a sequence of fragmentary brief looks and have a judicious focus on relevant sections to better understand the structure of vision [26].

a) Channel attention:

The channel attention map was created to benefit from interactions between features across channels by considering each channel of a feature map as a feature detector. The main objective of channel attention is to record the significant elements that contribute to the interpretation of an incoming image [31]. Architecturally, channel attention is calculated by shrinking the input feature map's geographic extent and using average- and maximum-pooled features to aggregate spatial data. Empirical studies have shown that the integration of these features considerably improves the cognitive capacity of networks

b) Spatial Attention:

Inter-spatial interactions between features allowed for the creation of a spatial attention map. Spatial attention, which works in conjunction with channel attention, lays greater emphasis on "where," a component of information than channel attention does.To generate informative feature descriptors from the channel axis, one commonly employs average and max pooling operations. These pooling operations aim to condense the information across the channel axis, thereby extracting meaningful representations. It is shown how effectively informative regions are highlighted when pooling operations are appealed along the channel axis.

A one-dimensional channel attention map $CAF \in R^{C\times1\times1}$ and the two-dimensional spatial attention map as SAF $\in R^{1\times H\times W}$ are progressively obtained by CBAM given an intermediary feature map $F \epsilon R^{C\times H\times W}$ as input. The overall attention process can be described as follows:

$$F1 = CAF(F) \otimes F \quad (2)$$

$$F2 = SAF(F1) \otimes F1 \quad (3)$$

where the $\otimes$ symbol stands for multiplication by elements. Channel attention values are broadcast along the spatial dimension, and vice versa, accordance with the

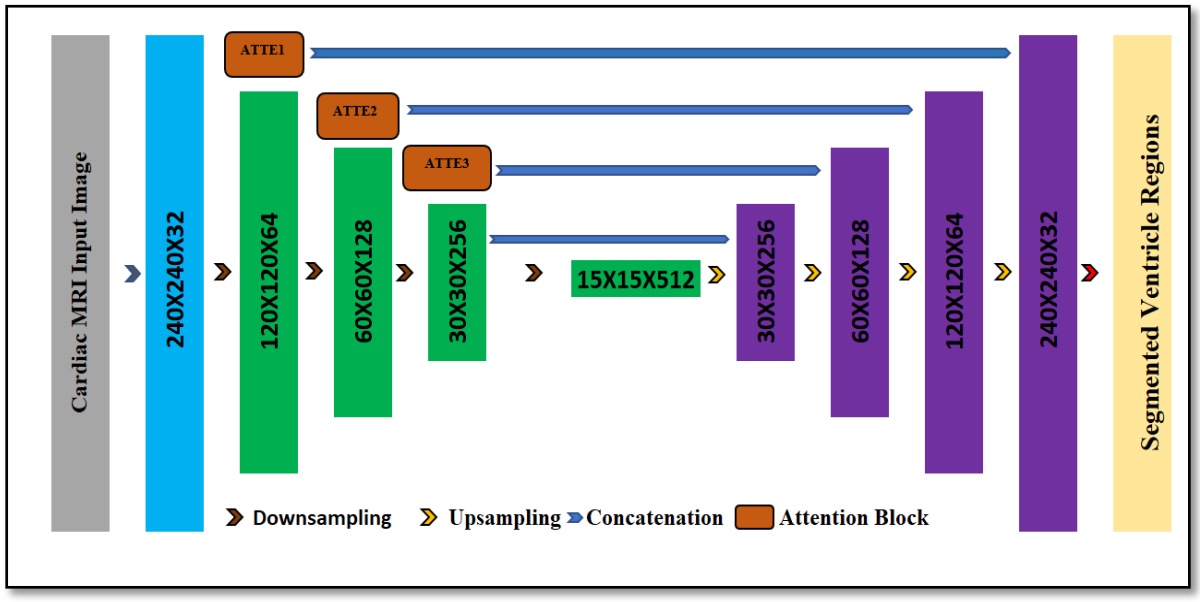


Figure 6: Detailed Architecture of UNet-CBAM

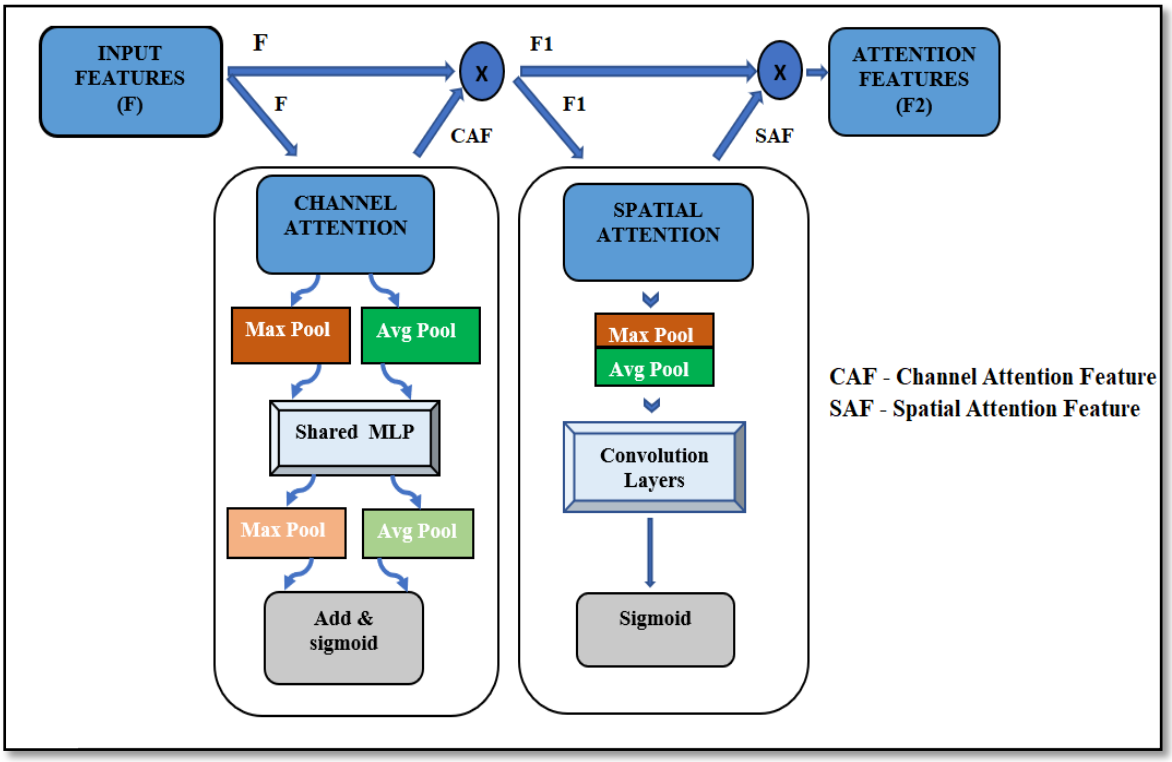


Figure 7.Channel and spatial submodules of the attention subsystem- ATTE1, ATTE2, ATTE3

multiplication process, which dictates how the attention values are broadcast (copied). Illustrations of each attention submodule are shown in Figure 7: Using a pooled network, the channel sub-module takes advantage of both maximum and average pooling outputs; The spatial sub-module sends identical two outputs to a convolution layer after pooling them along the channel axis.

Fc avg and Fc max, which stand for average and max-pooled features, respectively, are two unique spatial context descriptors that are generated in order to extract the spatial data. After that, a shared network receives both descriptors to create the channel attention map, CAF∈R.The shared network consists of multi-layer perceptrons (MLP) with a single hidden layer. Use element-wise aggregation to combine the resulting feature vectors once each descriptor has been subjected to the common network. The channel attention is established as follows:

$$CAF(F)=\sigma(MLP\ (AP(F))+ MLP(MP(F))) \qquad (4)$$

where σ represents the sigmoid function.Two pooling methods are used to integrate the channel information of a feature map to create two 2D spatial attention maps $F^s_{avg} \in R^{1\times H\times W}$ and $F^s_{max} \in R^{1\times H\times W}$. Average pooling, which computes the average value of each channel over the spatial dimensions, is the initial operation. The second procedure, known as max pooling, determines each channel's highest value across all spatial dimensions. The computation for spatial attention is as follows.

$$SAF(F)=\sigma(f([AP(F);MP(F)])) \qquad (5)$$

Here, f stands for a convolution operation and σ symbolises the sigmoid function. Figure 3 illustrates the segmentation output using various methods for the diastolic Phase. Figure 8 illustrates the segmentation output using various methods for the diastolic Phase.

4. EXPERIMENTAL SETUP

A. Performance Measures

The Dice coefficient and Hausdorff distance are used to determine the success rate of the various segmentation strategies previously discussed. For the quantitative analysis of these processes, precision, recall, F-score, and accuracy are the metrics employed. The AMD Ryzen system setup is used to implement this task in Python.

B.Dice Coefficient and Hausdorff Distance

The dice coefficient is a reliable metric for assessing the pixel-level similarity between a predicted segmentation and the corresponding real-world data.

$$DC = \frac{2AP}{2AP+IN+IP} \quad (6)$$

HD is utilized to compare ground truth images with segmentation results, enabling the ranking of different segmentation outcomes.

$$HD(P,Q)=max\{h(P,Q),h(P,Q)\} \quad (7)$$

Where P represents segmented output and Q signifies ground truth.

C. Precision and Recall

Precision can be expressed as the percentage of actual occurrences among all instances that have been retrieved. Recall is the proportion of occurrences that can be retrieved out of all relevant instances.

$$Precision\ (Pr) = \frac{AP}{AP+IP} \quad (8)$$

$$Recall\ (Re) = \frac{AP}{AP+IN} \quad (9)$$

D. F1Score

The F1 Score is measured by combining precision and recall.

$$F1\ Score = \frac{(2 * Pr * Re)}{(Pr+Re)} \quad (10)$$

Here, AP –AccuratePositive, IP–Incorrect Positiveand, IN–Incorrect Negative.

E. Dataset

ACDC (Automated Cardiac Diagnosis Challenge) dataset:This initiated effort makes use of the ACDC dataset. This dataset, which includes data from 150 individual patients, was produced using actual clinical examinations at the University Hospital of Dijon. This dataset has enough examples to cover many identified diseases. One healthy subject group and four diseased subgroups make up this dataset's five evenly distributed subgroups. A total of 100 patients' systole and diastole stages of cardiac MRI data are included; 80% of these data are erratically assigned to training, while the left 20% are used for testing.

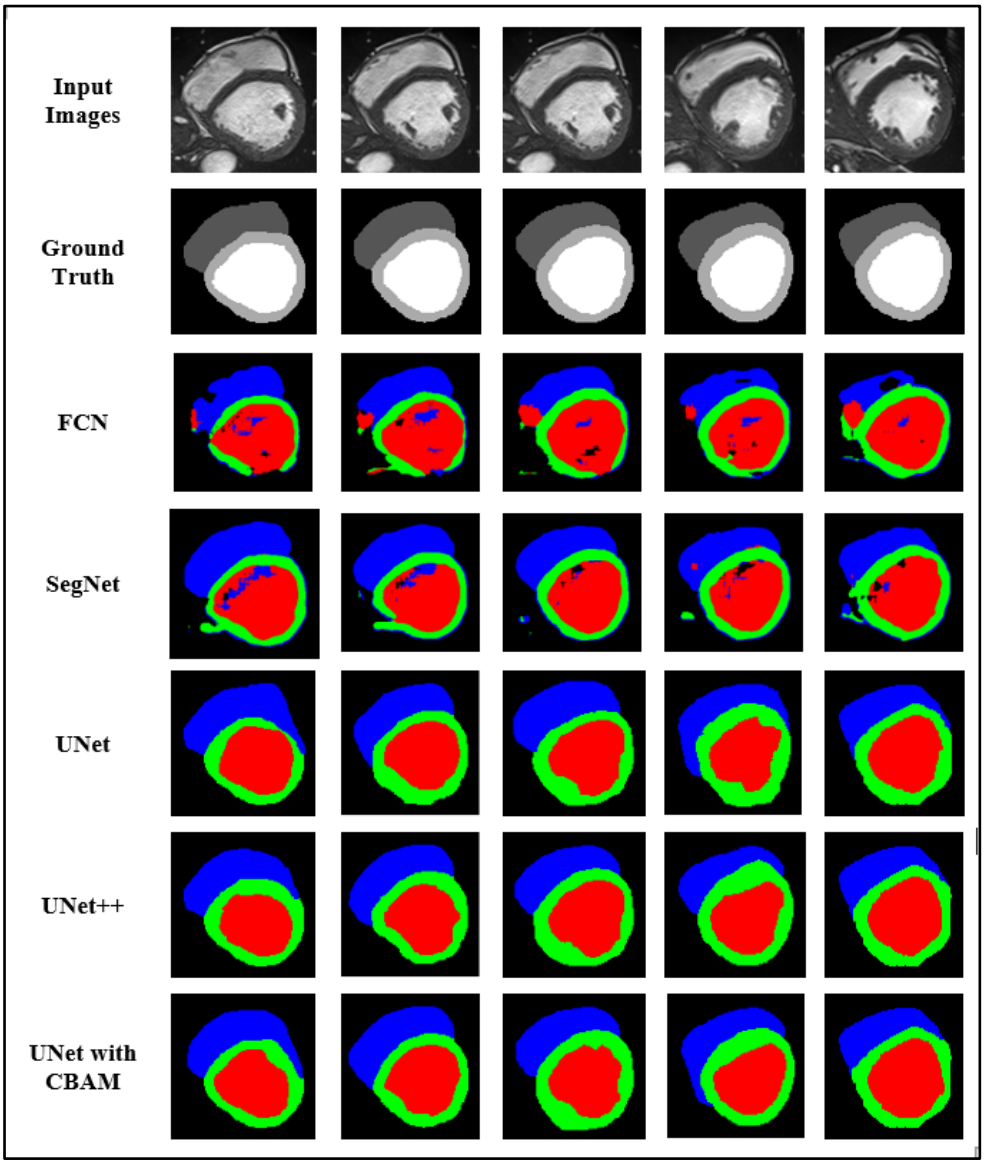


FIGURE 8:The Comparative Results for Segmented Ventricle Regions in Diastole Phase

5 . Experimental Results

The various cardiac image segmentation techniques were discussed in the previous sections, and this section evaluates the performance of these techniques.

A.Performance Analysis of Various Deep Learning Segmentation Techniques

Various image segmentation techniques as discussed in section 3 are employed for cardiac image segmentation. Both the systolic and diastolic stages of the process result in the division of the LV, RV, and MYO regions. When the heart contracts, it is in systole and when it relaxes thereafter, it is in diastole. By extracting these regions is very much useful in the identification of heart diseases. The performance of these techniques based on DC and HD is tabulated in Table 1. The experiment is carried out for different epochs like 20,40,60,80,100 and 120 and the experimental outputs are noted.

Table 1 shows that in epoch 20, the DC for FCN reached values of 0.607 for LV, 0.541 for RV, and 0.554 for MYO. Due to the lack of spatial consistency, FCN achieves less DC than SegNet, UNet, UNet++, and UNet-CBAM. For LV, RV, and MYO, respectively, the maximum value of FCN for epoch 120 is 0.835, 0.756, and

0.774.Table 1 makes it abundantly clear that when the epoch values rise, the DC values increase and the HD values decrease, and vice versa. UNet gained 0.03 more than SegNet because of its use of global location and context at the same time.UNet++ obtained better than UNet in terms of DC advantage due to its dense skip connections and convolution layers on skip paths, which work to bridge the semantic divide between decoder and encoder feature maps. As a result, UNet's performance is surpassed. A combination of UNet and CBAM was attempted, and the results were superior to those of all other segmentation methods covered in this study in terms of performance. UNet will perform better when the attention modes are combined. For LV, RV, and MYO, UNet-CBAM achieved DC as 0.93, 0.894, and 0.918, respectively.

Table 1: Quantitative Assessment of Segmentation Methods: DC and HD

Methods	Epochs	Systole						Diastole					
		DC			HD			DC			HD		
		LV	RV	MYO	LV	RV	MYO	LV	RV	MYO	LV	RV	MYO
FCN	20	0.587	0.567	0.569	22.4	25.6	20.9	0.607	0.541	0.554	22.8	26.1	20.8
	40	0.643	0.623	0.609	18.7	23.9	18.1	0.657	0.569	0.576	22.1	25.3	18.7
	60	0.704	0.689	0.645	13.6	22.5	16.3	0.708	0.601	0.617	19	24.1	16.8
	80	0.756	0.743	0.726	12.6	20.3	14.7	0.756	0.668	0.654	18.1	23.8	15.3
	100	0.806	0.777	0.812	11.7	19.1	13.2	0.798	0.708	0.723	16.4	21.4	14.6
	120	0.861	0.834	0.856	10.1	17.3	12.4	0.835	0.756	0.774	15.6	19.3	13.1
SegNet	20	0.601	0.597	0.586	21.7	23.8	20.7	0.654	0.652	0.604	14.9	16.3	17.2
	40	0.653	0.643	0.634	18.9	22.4	19.6	0.701	0.703	0.653	12.2	15.3	16.2
	60	0.726	0.713	0.687	16.7	21.5	18.6	0.751	0.756	0.707	11.6	14.9	14.5
	80	0.786	0.759	0.745	14.6	20.4	15.7	0.842	0.803	0.765	10.3	13.2	13.1
	100	0.835	0.802	0.823	13.1	18.6	13.4	0.876	0.836	0.829	9.1	12.1	11.4
	120	0.903	0.862	0.885	8.4	16.7	11.9	0.927	0.915	0.875	8.4	11.8	10.8
UNet	20	0.621	0.607	0.615	19.8	21.6	20.5	0.687	0.659	0.668	18.1	18.4	16.9
	40	0.684	0.658	0.667	17.6	20.1	19.8	0.734	0.729	0.705	16.9	17.6	15.6
	60	0.748	0.721	0.736	16.2	18.5	17.3	0.786	0.805	0.756	14.7	16.7	13.7
	80	0.814	0.785	0.794	14.9	16.7	15.7	0.856	0.841	0.806	11.5	14.3	12.3
	100	0.846	0.824	0.837	12.7	15.3	14.4	0.887	0.876	0.849	9.5	13.4	11.3
	120	0.917	0.886	0.903	9.5	14.8	10.2	0.941	0.933	0.897	6.7	12.7	8.4
UNet++	20	0.652	0.628	0.633	18.6	20.1	19.1	0.725	0.709	0.687	17.1	19.5	20.6
	40	0.705	0.683	0.697	16.8	19.5	18.2	0.801	0.768	0.728	15.9	18.5	19.1
	60	0.761	0.735	0.746	14.3	17.6	16.5	0.865	0.807	0.785	13.7	16.1	15.9
	80	0.836	0.801	0.824	12.9	15.2	14.4	0.889	0.867	0.836	11.1	14.9	13.8
	100	0.877	0.854	0.866	11.7	14.4	12.7	0.924	0.908	0.864	9.1	13.8	12.7
	120	0.924	0.889	0.911	9.4	13.7	11.9	0.946	0.938	0.898	6.6	11.9	8.3
UNet-CBAM	20	0.665	0.632	0.649	16.7	19.4	18.1	0.741	0.718	0.745	17.8	20.5	23.7
	40	0.723	0.707	0.711	15.2	18.1	17.7	0.761	0.847	0.769	15.6	18.4	18.5
	60	0.788	0.751	0.767	13.8	16.2	15.4	0.821	0.863	0.843	14.5	15.6	15.6
	80	0.841	0.823	0.834	11.5	14.6	13.2	0.911	0.876	0.865	11.3	14.5	13.4
	100	0.896	0.877	0.886	9.4	13.9	11.6	0.948	0.929	0.854	8.9	12.3	10.1
	120	0.93	0.894	0.918	7.6	12.4	11.8	0.95	0.941	0.899	6.5	10.5	8.3

Addressing HD in both the systole and the diastole phases Performance-wise, UNet outperforms SegNet and FCN. All other techniques, including FCN, SegNet, UNet, and UNet++, are outperformed by the UNet-CBAM.Next, the performance of these techniques is evaluated based on Precision and Recall and the outcomes are tabulated in Table 2. For this experiment, the segmentation techniques like FCN, SegNet, UNet, UNet++ and UNet-CBAM are employed and the results are noted for different epochs. FCN achieved an average of 0.84 precision, 0.81 recall for the systole phase and 0.84 precision and 0.81 recall for the diastole phase at epoch 120. At epoch 120SegNetachieved an average of 0.86 precision, 0.85 recall for the systole phase and 0.89 precision

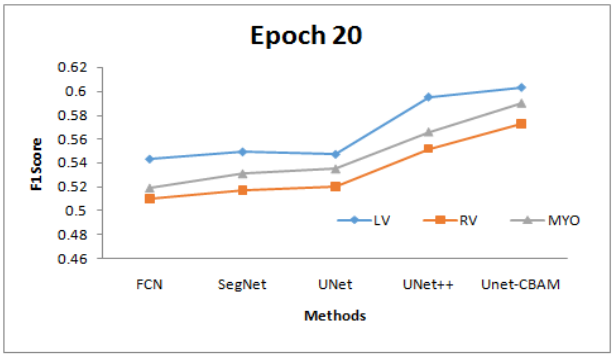
and 0.87 recall for the diastole phase which has a gain of 0.01 to 0.03 over FCN. At epoch 120 UNet++ achieved an average of 0.89 precision, and 0.88 recall for the systole phase which is more than UNet. Overall UNet-CBAM has attained an average of 0.90 precision and 0.89 recall at epoch 120 which is higher than all other techniques.

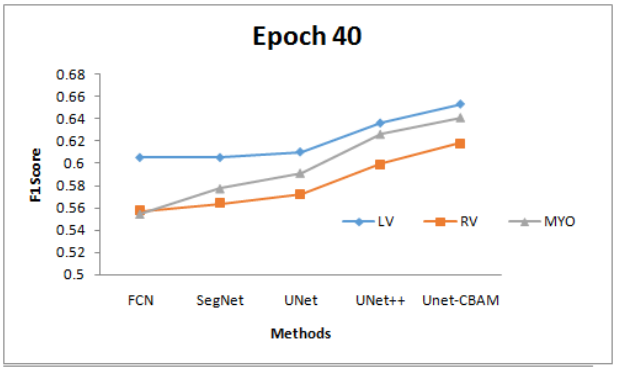
Table 2 :Performance analysis of various segmentation techniques based on Precision and Recall

Methods	**Epchos**	**Systole**						**Diastole**					
		Pr			**Re**			**Pr**			**Re**		
		LV	**RV**	**MYO**	**LV**	**RV**	**MYO**	**LV**	**RV**	**MYO**	**LV**	**RV**	**MYO**
FCN	**20**	0.521	0.512	0.529	0.567	0.509	0.509	0.607	0.584	0.589	0.624	0.612	0.573
	40	0.602	0.557	0.552	0.609	0.558	0.556	0.645	0.621	0.615	0.664	0.621	0.615
	60	0.635	0.604	0.612	0.657	0.605	0.612	0.702	0.687	0.652	0.703	0.654	0.652
	80	0.712	0.687	0.698	0.703	0.678	0.678	0.746	0.721	0.702	0.751	0.708	0.684
	100	0.801	0.765	0.778	0.756	0.756	0.765	0.809	0.805	0.756	0.79	0.765	0.743
	120	0.845	0.839	0.854	0.829	0.805	0.821	0.854	0.845	0.835	0.836	0.802	0.789
SegNet	**20**	0.543	0.507	0.523	0.556	0.528	0.516	0.623	0.602	0.601	0.645	0.624	0.653
	40	0.612	0.564	0.587	0.598	0.564	0.567	0.697	0.653	0.645	0.686	0.675	0.702
	60	0.645	0.623	0.643	0.652	0.612	0.623	0.753	0.708	0.678	0.756	0.729	0.745
	80	0.723	0.705	0.712	0.724	0.702	0.709	0.802	0.756	0.758	0.816	0.786	0.786
	100	0.802	0.786	0.803	0.803	0.765	0.796	0.865	0.821	0.846	0.867	0.823	0.834
	120	0.864	0.856	0.873	0.861	0.837	0.859	0.893	0.871	0.882	0.884	0.876	0.871
UNet	**20**	0.554	0.524	0.547	0.541	0.517	0.523	0.676	0.612	0.645	0.665	0.634	0.619
	40	0.616	0.571	0.594	0.604	0.573	0.589	0.712	0.689	0.669	0.709	0.689	0.667
	60	0.671	0.645	0.663	0.666	0.636	0.654	0.789	0.768	0.705	0.745	0.743	0.708
	80	0.757	0.729	0.736	0.735	0.714	0.726	0.82	0.821	0.791	0.785	0.795	0.765

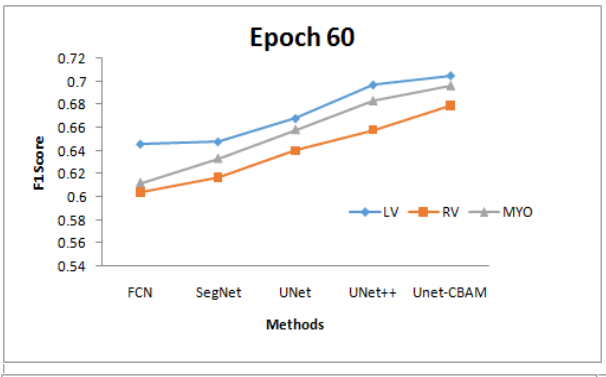
	100	0.822	0.806	0.812	0.816	0.791	0.807	0.881	0.886	0.865	0.851	0.845	0.846
	120	0.897	0.883	0.891	0.887	0.865	0.874	0.911	0.893	0.901	0.903	0.894	0.907
UNet++	**20**	0.609	0.557	0.571	0.581	0.547	0.561	0.694	0.643	0.669	0.687	0.62	0.667
	40	0.643	0.605	0.638	0.629	0.593	0.614	0.745	0.701	0.708	0.729	0.679	0.715
	60	0.709	0.664	0.689	0.686	0.652	0.677	0.804	0.756	0.768	0.765	0.726	0.768
	80	0.761	0.733	0.746	0.744	0.724	0.739	0.867	0.806	0.825	0.804	0.789	0.821
	100	0.826	0.816	0.819	0.822	0.806	0.814	0.883	0.867	0.876	0.861	0.834	0.889
	120	0.905	0.891	0.897	0.893	0.871	0.879	0.928	0.901	0.915	0.912	0.907	0.911
UNet-CBAM	**20**	0.614	0.578	0.609	0.592	0.569	0.572	0.711	0.734	0.729	0.716	0.716	0.711
	40	0.663	0.626	0.653	0.644	0.611	0.629	0.794	0.751	0.786	0.781	0.792	0.794
	60	0.714	0.684	0.705	0.697	0.675	0.687	0.853	0.845	0.856	0.854	0.841	0.853
	80	0.779	0.745	0.761	0.753	0.722	0.731	0.881	0.875	0.894	0.861	0.874	0.881
	100	0.845	0.827	0.831	0.836	0.817	0.824	0.892	0.919	0.916	0.906	0.895	0.892
	120	0.917	0.894	0.902	0.908	0.878	0.886	0.933	0.919	0.923	0.928	0.914	0.921

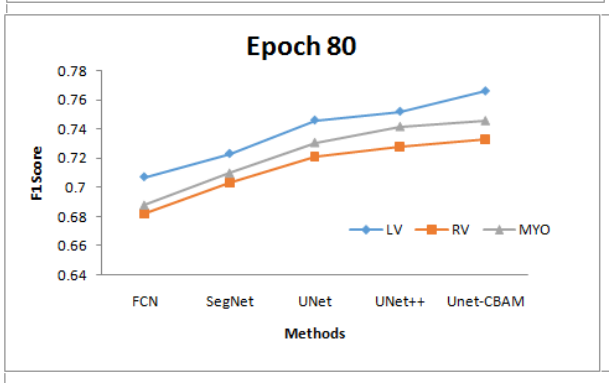
Figure 9 demonstrates how the F1 Score at various epochs in the systole phase is used to evaluate the efficacy of these methods. For each epoch, UNet CBAM has outperformed alternative approaches in terms of results.


Epoch 20
F1Score
0.62
0.6
0.58
0.56
0.54
0.52
0.5
0.48
0.46
LV
RV
MYO
FCN
SegNet
UNet
UNet++
Unet-CBAM
Methods

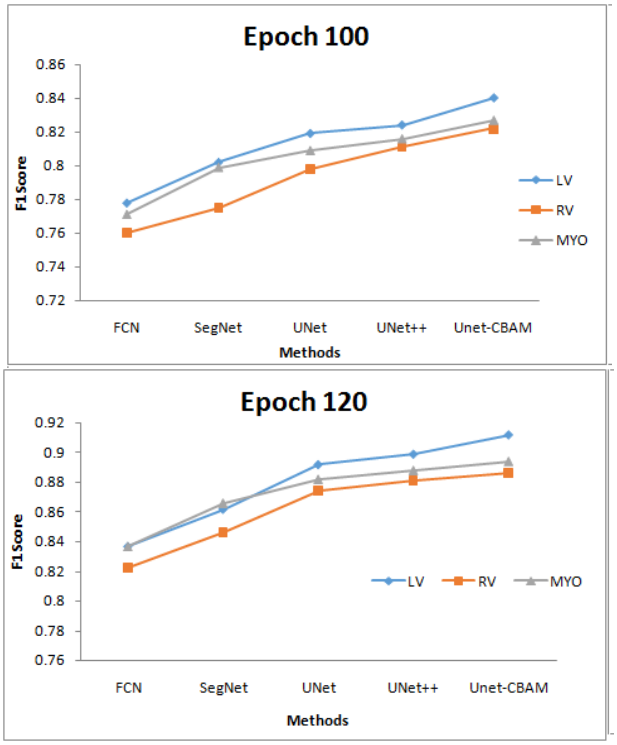

Epoch 40
F1Score
0.68
0.66
0.64
0.62
0.6
0.58
0.56
0.54
0.52
0.5
LV
RV
MYO
FCN
SegNet
UNet
UNet++
Unet-CBAM
Methods

(a) (b)


Epoch 60
F1Score
0.72
0.7
0.68
0.66
0.64
0.62
0.6
0.58
0.56
0.54
LV
RV
MYO
FCN
SegNet
UNet
UNet++
Unet-CBAM
Methods


Epoch 80
F1Score
0.78
0.76
0.74
0.72
0.7
0.68
0.66
0.64
LV
RV
MYO
FCN
SegNet
UNet
UNet++
Unet-CBAM
Methods

(c) (d)



(e) (f)

Figure 9. F1 Score achieved for various techniques at different epochs in the systole phase for ACDC dataset

The F1 score for the systole phase at various epochs is shown in Figure 9. This figure makes it clear that the UNet-CBAM produce good results when compared to other techniques.

B. Comparison of Training Timefor Various Segmentation Techniques

Table 3 provides information on the training times (in seconds) of different deep-learning methods for the segmentation of the phases of the cardiac cycle in the ACDC dataset. The methods evaluated in this study include FCN, SegNet, UNet, UNet++, and UNet-CBAM. The results show that UNet had the shortest training time for both the diastolic and systolic phases, with 299 seconds for the diastolic and 244 seconds for the systolic phases. UNet was closely followed by SegNet, which had a training time of 350 seconds for diastolic and 320 seconds for systolic phases. UNet++ and SegNet had similar training times of 337 seconds and 312 seconds, respectively, for the diastolic and systolic phases.

Table 3: Performance comparison of various segmentation techniques based on Training Time

Time taken for Training(sec)		
Techniques	**Diastolic**	**Systolic**
FCN[18]	486	414
SegNet[19]	354	326
UNet[20]	297	247
UNet++[21]	340	315
UNet-CBAM [22]	396	325

FCN had the longest training time of all the methods evaluated, with 410 sec for the diastolic phase and 490 sec for the systolic phase. UNet-CBAM also had a longer training time compared to the other methods, with 394 seconds for the diastolic and 323 sec for the systolic.Overall, the results suggest that UNet has the shortest training time among the evaluated methods.

6. Conclusion

Deep learning techniques have shown to be very important in the field of image segmentation, and they are predicted to get better with further study and development. The effectiveness of the segmented images is trustworthy for clinical monitoring. Overall, UNet-CBAM has shown to be effective in improving segmentation results and usage is expected to continue to increase in the field of computer vision. However further research is needed to explore the full potential of attention mechanisms in medical imaging and to investigate their performance in more complex and diverse datasets.

References

[1] J. Habetha, "The MyHeart project-fighting cardiovascular diseases by prevention and early diagnosis." In "2006 International Conference of the IEEE Engineering in Medicine and Biology Society," pp. 6746-6749. IEEE, 2006.

[2] D. S. Celermajer, C. K. Chow, E. Marijon, N. M. Anstey, and K. S. Woo, "Cardiovascular disease in the developing world: prevalences, patterns, and the potential of early disease detection." "Journal of the American College of Cardiology," vol. 60, no. 14, pp. 1207-1216, 2012.

[3] M. C. Azad, W. D. Shoesmith, M. Al Mamun, A. F. Abdullah, D. K. S. Naing, M. Phanindranath, and T. C. Turin, "Cardiovascular diseases among patients with schizophrenia." Asian Journal of Psychiatry, vol. 19, pp. 28-36, 2016.

[4] D. Zhao, "Epidemiological features of cardiovascular disease in Asia." "JACC: Asia," vol. 1, no. 1, pp. 1-13, 2021.

[5] P. M. Elliott, A. Anastasakis, M. A. Borger, M. Borggrefe, F. Cecchi, P. Charron, A. A. Hagege, A. Lafont, G. Limongelli et al., "2014 ESC guidelines on diagnosis and management of hypertrophic cardiomyopathy the task force for the diagnosis and management of hypertrophic cardiomyopathy of the European society of cardiology (ESC)," "Eur. Heart J.," vol. 35, no. 39, pp. 2733–2779, 2014.

[6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," "Nature," vol. 521, no. 7553, pp. 436–444, 2015. https://doi.org/10.1038/nature14539.

[7] A. Anaya-Isaza, L. Mera-Jiménez, and M. Zequera-Diaz, "An overview of deep learning in medical imaging." "Informatics in Medicine Unlocked," vol. 26, 2021.

[8] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert et al., "nnu-net: Selfadapting framework for u-net-based medical image segmentation," arXiv preprint arXiv:1809.10486,2018.

[9] T. Wang, J. Xiong, X. Xu, M. Jiang, H. Yuan, M. Huang, J. Zhuang, and Y. Shi, "Msu-net: Multiscale statistical u-net for real-time 3d cardiac MRI video segmentation." In International Conference on Medical Image Computing and Computer-Assisted Intervention,pp. 614–622. Springer, 2019.

[10] T. Wang, X. Xu, J. Xiong, Q. Jia, H. Yuan, M. Huang, J. Zhuang, and Y. Shi, "Icaunet: Ica inspired statistical unet for real-time 3d cardiac cine MRI segmentation." In International Conference on Medical Image Computing and Computer-Assisted Intervention,pp. 447–457. Springer, 2020.

[11] S. Queir´os, D. Barbosa, B. Heyde, P. Morais, J. L. Vilac¸a, D. Friboulet, O. Bernard, and J. Dhooge, "Fast automatic myocardial segmentation in 4D cine CMR datasets." "Med. Image Anal.," vol. 18, no. 7, pp. 1115 – 1131, 2014.

[12] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-unet: Unet-like pure transformer for medical image segmentation." "arXiv preprint arXiv:2105.05537," 2021.

[13] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," "IEEE Transactions on Pattern Analysis and Machine Intelligence," vol. 40, no. 4, pp. 834-848, 2018.

[14] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "Ce-net: Context encoder network for 2d medical image segmentation," "IEEE Transactions on Medical Imaging," vol. 38, 2019.

[15] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network." In "2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)," 2017, pp. 6230-6239.

[16] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert, "Attention gated networks: Learning to leverage salient regions in medical images." "Medical Image Analysis," vol. 53, pp. 197-207, 2019.

[17] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks." In "2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition," 2018, pp. 7794-7803.

[18] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation." In "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition," pp. 3431-3440, 2015.

[19] 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.

[20] 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," pp. 234-241. Springer, Cham, 2015.

[21] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation." "IEEE Transactions on Medical Imaging," vol. 39, no. 6, pp. 1856-1867, 2019.

[22] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module." In "Proceedings of the European Conference on Computer Vision (ECCV)," pp. 3-19, 2018.

[23] G. Gomathi, V. Subha, "Semantic Segmentation of Ventricular and Myocardium Regions in Cardiac MRI," "Design Engineering," ISSN: 0011-9342, Issue: 9, Pages: 8510-8522, 2021.

[24] H. Wang, F. Shi, L. Wang, S.-C. Hung, M.-H. Chen, S. Wang, W. Lin, and D. Shen, "Dilated dense U-Net for infant hippocampus subfield segmentation." "Frontiers in neuroinformatics," vol. 13, 2019, 30.