life

Article

Fully Automatic Left Ventricle Segmentation Using Bilateral Lightweight Deep Neural Network

Muhammad Ali Shoaib [1,2], Joon Huang Chuah [1], Raza Ali [1,2], Samiappan Dhanalakshmi [3], Yan Chai Hum [4], Azira Khalil [5] and Khin Wee Lai [6,*]

[1] Department of Electrical Engineering, Faculty of Engineering, Universiti Malaya, Kuala Lumpur 50603, Malaysia
[2] Faculty of Information and Communication Technology, BUITEMS, Quetta 87300, Pakistan
[3] Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, Kattankulathur 603203, India
[4] Department of Mechatronics and Biomedical Engineering (DMBE), Lee Kong Chian Faculty of Engineering and Science (LKC FES), Universiti Tunku Abdul Rahman (UTAR), Jalan Sungai Long, Bandar Sungai Long, Cheras, Kajang 43000, Malaysia
[5] Faculty of Science and Technology, Universiti Sains Islam Malaysia (USIM), Nilai 71800, Malaysia
[6] Department of Biomedical Engineering, Faculty of Engineering, Universiti Malaya, Kuala Lumpur 50603, Malaysia
* Correspondence: lai.khinwee@um.edu.my

Abstract: The segmentation of the left ventricle (LV) is one of the fundamental procedures that must be performed to obtain quantitative measures of the heart, such as its volume, area, and ejection fraction. In clinical practice, the delineation of LV is still often conducted semi-automatically, leaving it open to operator subjectivity. The automatic LV segmentation from echocardiography images is a challenging task due to poorly defined boundaries and operator dependency. Recent research has demonstrated that deep learning has the capability to employ the segmentation process automatically. However, the well-known state-of-the-art segmentation models still lack in terms of accuracy and speed. This study aims to develop a single-stage lightweight segmentation model that precisely and rapidly segments the LV from 2D echocardiography images. In this research, a backbone network is used to acquire both low-level and high-level features. Two parallel blocks, known as the spatial feature unit and the channel feature unit, are employed for the enhancement and improvement of these features. The refined features are merged by an integrated unit to segment the LV. The performance of the model and the time taken to segment the LV are compared to other established segmentation models, DeepLab, FCN, and Mask RCNN. The model achieved the highest values of the dice similarity index (0.9446), intersection over union (0.8445), and accuracy (0.9742). The evaluation metrics and processing time demonstrate that the proposed model not only provides superior quantitative results but also trains and segments the LV in less time, indicating its improved performance over competing segmentation models.

Keywords: left ventricle; deep learning; spatial features; channel features

check for updates

Citation: Shoaib, M.A.; Chuah, J.H.; Ali, R.; Dhanalakshmi, S.; Hum, Y.C.; Khalil, A.; Lai, K.W. Fully Automatic Left Ventricle Segmentation Using Bilateral Lightweight Deep Neural Network. *Life* **2023**, *13*, 124. https://doi.org/10.3390/life13010124

Academic Editor: Marcin Balcerzyk

Received: 1 December 2022
Revised: 22 December 2022
Accepted: 29 December 2022
Published: 1 January 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

According to the World Health Organization, cardiovascular diseases (CVDs) are one of the leading causes of death in the world, claiming over 17.9 million lives each year and accounting for approximately 31% of global fatalities [1]. As one of the leading causes of death, cardiovascular disease has drawn considerable attention in medical practice. To assess diagnostic features of the heart, such as ejection fraction and myocardial mass, it is necessary to measure the heart's volumes. Whereas the precise measurement of the left ventricle (LV) is the most important for figuring out these parameters and diagnosing cardiovascular diseases [2].

Imaging systems and other diagnostic devices such as echocardiography, angiography, magnetic resonance imaging, etc., [3] provide a reliable diagnosis, which is the first step in the control and treatment of CVD. Among these imaging techniques, echocardiography is unquestionably the preferred method for evaluating the heart ventricles. Echocardiography is a user-friendly, reliable, and non-invasive method [4]. It has significant uses in diagnosis and decision-making for numerous CVD. As a result, the echocardiography approach is now the most commonly used approach for evaluating and assessing patients with cardiovascular disease [5,6].

The automatic LV segmentation from echocardiography images can help the sonographer in the detection of diseases. However, semi-automatic or manual annotation is still a part of daily clinical practice. This results in tedious jobs susceptible to intra- and inter-observer variation [7]. In addition, the intrinsic challenges in segmenting echocardiography images, such as low contrast, brightness inhomogeneities, and changes in the speckle pattern, are well established, and several methods [8] are required to eliminate noise to improve the visibility of anatomical features in images.

Recent years have seen the emergence of DL as a fully automated image processing method. The success of the DL model in the medical sector has resulted in a surge of data mining, and the use of DL in a variety of medical areas has particularly gained interest in the field of medical image analysis [9–11]. DL has begun to incorporate image registration, detection of the lesion, study of brain function, and image segmentation [12–14].

This study, motivated by the effectiveness of DL techniques, proposed a DL technique for LV segmentation from echocardiography images. The following are the main contributions of the proposed method:

- An innovative and lightweight single-stage instant segmentation DL method for fast and accurate LV segmentation from echocardiography images.
- The spatial and contextual data are derived from a single, lightweight backbone model, making the procedure fast.
- The spatial and contextual features are refined simultaneously by two distinct units, which improve the performance of the model.
- An integration unit is used to efficiently combine the features.

The remaining sections of this work are structured as follows: Section 2 provides an overview of the existing LV segmentation architectures. The explanation of each module of the proposed methodology is outlined in Section 3. The experimental setup, datasets, and evaluation matrices are described in Section 4 of the paper. The findings and discussion are presented in Section 5.

2. Literature Review

Various techniques, such as the active appearance model, deformable models, and machine learning, have been utilized for LV segmentation and have produced satisfactory results [15–17]. Most of these techniques rely on handcrafted features, which may amount to an overly simplistic source of information. In contrast, deep learning (DL) has shown considerable progress in recent studies from the perspective of medical image segmentation techniques [18]. DL models, such as fully convolutional networks (FCN) [19], and U-Net [20], can be referred to as single-stage segmentation model and have attained enormous success in the image segmentation task. The encoder–decoder architectures are employed by both FCN and U-Net. The features are extracted by the encoder, and classification and regression of pixels are performed using those features, whereas the decoder is intended for upsampling to rebuild a semantic segmentation mask according to input image size.

For LV segmentation in CT images, an FCN-based architecture is utilized with pre-trained weights of VGG [21]. Another revised version of FCN with different loss functions is analyzed in [22] and an iterative multi-path FCN (IMFCN) segmentation model is proposed, which segments the LV and RV from MRI images. However, the most significant weakness of FCN is the loss of spatial information. Likewise, U-Net is another widely used DL architecture for LV segmentation [23,24]. The anatomically constrained neural

network proposed in [25] is yet another U-Net-based architecture that is employed for LV segmentation from 3D images. The effectiveness of the DL network and the quantity of data needed to train the model for the segmentation of LV by applying the U-Net model were examined by Leclerc et al. [26,27].

In the literature, a few one-stage attention models for medical image segmentation have been proposed [28–30]. A model using the attention gate that computes the attention coefficients that scale the input characteristics was proposed by [31]. On top of the U-Net model, a segmentation model with attention gates is developed, and attention gates are transmitted by skip connections on the decoding part of the U-Net. Also, the suggested approach in [32] is based on the attention U-Net segmentation model. An input image pyramid and deeply supervised output are both integrated into the attention U-Net model. This approach segments the desired region and avoids extracting the same low-level characteristics again and again. In [33], a lightweight segmentation model is designed for LV segmentation from echocardiography images. The proposed approach provides a more optimal balance between the number of parameters and segmentation performance. In another study on pediatric echocardiography images, the LV and the left atrium (LA) are segmented using an attention mechanism. The purpose of a spatial path is to extract spatial characteristics, whereas the function of a contextual path is to extract contextual features. Utilizing the fusion model, these two modules' characteristics are combined [34]. A CNN-based model extracts the spatial data, whereas a backbone model retrieves the contextual information. Although this attention-based LV segmentation achieves high segmentation performance, it may require more time to train and segment since it extracts spatial data using a CNN model and context using another backbone model.

Mask R-CNN [35,36], on the other hand, is a well-known two-stage segmentation model that has achieved an immense amount of success in image segmentation. The CNN is used in the first stage of the two-stage model to filter out some of the proposal boxes, and then the bounding boxes are classified and regressed. The performance of the two-stage model is much better than that of the one-stage model. A comparative study [37] on the performance of Mask RCNN, FCN, and SegNet for LV segmentation indicated that Mask RCNN segmented the LV relatively better than FCN and SegNet [37].

A two-stage model for calculating the clinical indices by segmenting the LV is proposed in [38]. In this study, initially, the LV centers were detected across the cardiac cycle using a Res-circle network. Based on this information, new images with the LV cavity in focus were produced using a cropping approach with pre-defined fixed dimensions. Leclerc et al. [39] presented an attention mechanism intending to improve the segmentation in echocardiography images. The results from the first segmentation network are multiplied by a binary map of the attention model. The U-Net model receives this data as input and uses it to perform the final segmentation. Even though this model decreased the outliers, it did not enhance the accuracy overall. Based on the segmentation performance of the U-Net, [40] proposed a multistage LV segmentation model, referred to as a multistage attention model. In the first phase, U-Net is utilized to suggest the RPN, and in the second phase, regression and segmentation networks execute the segmentation of the LV from echocardiography images.

In summary, models based on convolutional neural networks (CNNs) have shown remarkable improvement over non-DL methods in LV segmentation. The single-stage models are fast enough, but when compared to the two-stage models, the single-stage models' speed comes at the expense of accuracy when predicting instance classes. Even though two-stage models have improved their level of accuracy, these models are still slow for real-time processing of LV segmentation.

3. Methodology

The proposed LV segmentation model is shown in Figure 1. The training images along with binary masks are transmitted to the backbone model and feature pyramid network, which extracts both low-level and high-level image features using the binary mask of

training images. The spatial feature and channel feature modules, respectively, process these low-level and high-level retrieved features. The output of these two modules is then processed by the feature integration unit, which outputs the segmented LV. The following sections explain the entire process in detail.

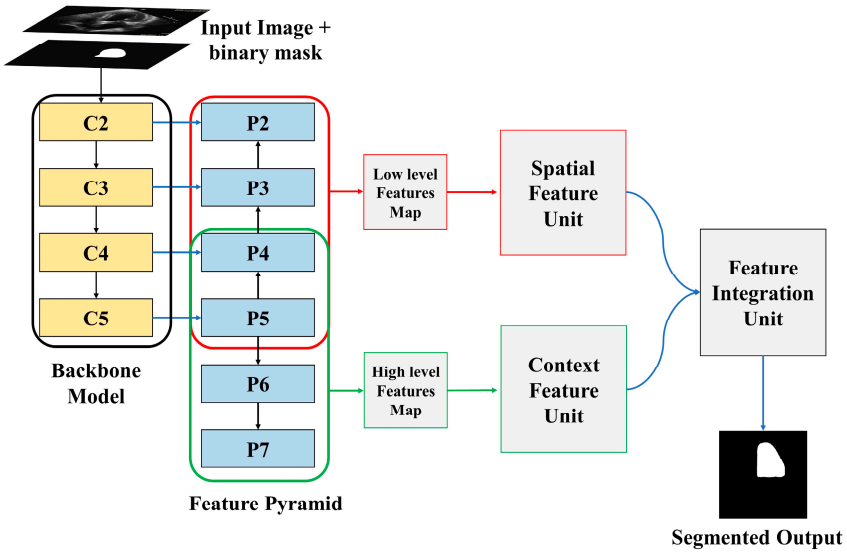


Figure 1. Overview of proposed model.

3.1. Backbone Model and Feature Pyramid Network

Residual networks, abbreviated as ResNet, is a traditional neural network used as the backbone for several computer vision applications. ResNet enables the training of hundreds or thousands of layers while still achieving impressive performance. ResNet 50 is used as a backbone model in this study. From a single-scale image of any size, a feature extractor called a feature pyramid network, or FPN, generates correspondingly scaled feature maps at several layers in a convolutional way. It produces a features pyramid that has a similar level of semantic weight regardless of the scale. This objective is simply achieved by building an additional pathway that incorporates high-level features from the upper layers to the lower layers with better resolution, but weaker semantic information.

The input image and binary masks passed through two different channels. The "bottom-up pathway" is the first, while the "top-down pathway" is the second. The spatial sizes must be consistently decreased from bottom to top to generate the pyramid feature maps at different stages of the pyramid. As the image along with the binary mask is passed from bottom to top, the spatial dimension decreases by a factor of one half, achieved by making the stride double. The outputs of the backbone model are denoted by C1, C2, C3, C4, and C5. The output of C5 is employed as the reference set of feature maps for strengthening the top-down path.

The semantically richer feature from higher pyramid levels is translated to higher spatial resolution by the top-down part of FPN by performing the nearest neighbor technique with a factor of 2. Using the skip connection, features of each lateral connection were subsequently merged with equivalent features from the bottom-up path. The bottom-up features undergo 1×1 convolution to reduce the channel dimensions, enabling the merging of bottom-up and top-down features. The element-wise addition is used to combine the feature maps from the top-down pathway and the bottom-up pathway. A 3×3 convolutional filter is subsequently applied to generate the final feature map. Two more pyramid levels, P6 and P7, are formed in addition to P2 to P5. P7 is created by applying a 3×3 filter to P6 whereas P6 is created by applying a 3×3 filter to the last convolution layer of the backbone model. The extracted features P2 to P5 are merged and referred to as low-level features, whereas the extracted features P5 to P7 are merged and referred to as high-level features.

3.2. Spatial Feature Unit

The spatial feature unit (SFU) concentrates on the "where" informative element of the image. The refined and rich spatial features are generated by leveraging the spatial relationship between the features. These spatial features can be obtained by applying a specific operation (P_x) which takes the input function or tensor $S_T \epsilon {S_T}^{H \times W \times C}$, where C is a channel plane with a spatial dimension of H x W, and produces a flattened 2D tensor (${S_T}^{H \times W}$) over the spatial dimension. In the proposed study the two pooling processes (maximum and average) along the channel axis are used to determine the refined spatial features.

$$P_a\left({S_T}^{H \times W \times C}\right) \rightarrow S_{Ta}^{H \times W} \tag{1}$$

$$P_m\left({S_T}^{H \times W \times C}\right) \rightarrow S_{Tm}^{H \times W} \tag{2}$$

where P_a and P_m in Equations (1) and (2) represent the average and max pooling respectively. To aggregate the channel information, the outputs of Equations (1) and (2) are concatenated. The merged feature descriptor is then fed into a convolutional layer, batch normalization, and the output is passed through sigmoid and transformed to yield the intermediate feature tensor I_T represented in Equation (3).

$$S_T^I = \varsigma\left[BN\left\{f\left(S_{Ta}^{H \times W} \odot S_{Tm}^{H \times W}\right)\right\}\right] \tag{3}$$

where ς is the sigmoid operator, BN is batch normalization, f is the convolution operation and $\odot$ represents the concatenation operation. The final refined spatial features S'_T is produced by multiplying the input feature, S_T, with intermediate spatial tensor S_T^I, as represented in Equation (4). Figure 2 illustrates the entire process.

$$S'_T = S_T \times S_T^I \tag{4}$$

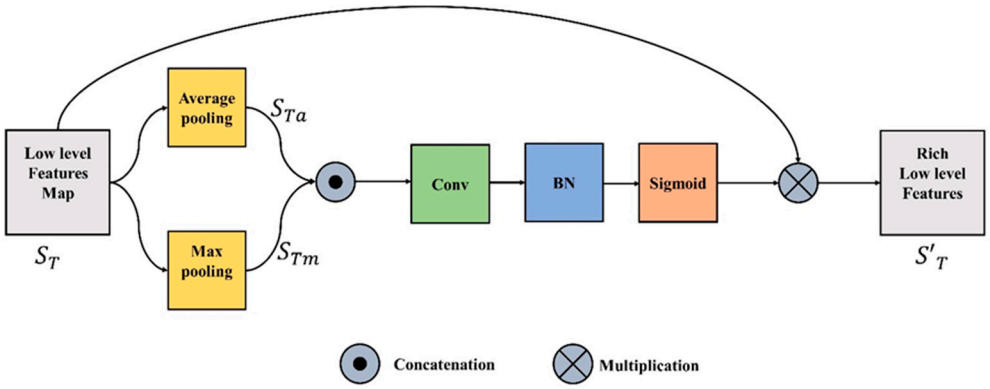


Figure 2. Spatial Feature Unit.

3.3. Channel Feature Unit

The focus of the channel features unit (CFU) is on "what" is significant in an input image since each channel of a feature map is thought of as a feature detector. The output of this unit is also known as channel attention and is attained by compressing the spatial dimension of the input feature map. Input channel feature or tensor $C_T \epsilon\ {C_T}^{H \times W \times C}$ is pooled using two different pooling operations, maximum and average pooling, unlike the spatial module, where the output is a 3D tensor as represented in Equations (5) and (6).

$$P_a\left({C_T}^{H \times W \times C}\right) \rightarrow C_{Ta}^{H \times W \times C} \tag{5}$$

$$P_m\left({C_T}^{H \times W \times C}\right) \rightarrow C_{Tm}^{H \times W \times C} \tag{6}$$

The maximum and average pooling not only successfully learns the extent of the desired object but also acquires crucial information about the object's specific characteristics to estimate finer channel-wise attention. The $C_{Ta}^{H\times W\times C}$ and $C_{Tm}^{H\times W\times C}$ are fed to a multi-layer perceptron with one hidden layer. This MLP is simply a fully connected layer and the output generated is shown in Equations (7) and (8).

$$\mathrm{MLP}\left({C_T}^{H\times W\times C}\right) = C_{Ta}^{1\times 1\times C} \tag{7}$$

$$MLP\left({C_T}^{H\times W\times C}\right) = C_{Tm}^{1\times 1\times C} \tag{8}$$

These two outputs are added up element by element, the reshaped summed tensor and sigmoid are applied on the output to generate the intermediate channel feature tensor C_T^I represented in Equation (9).

$$C_T^I = \varsigma\left[reshape\left(C_{Ta}^{1\times 1\times C} + C_{Tm}^{1\times 1\times C}\right)\right] \tag{9}$$

Equation (10) illustrates that the refined channel feature C_T' is produced by multiplying the input feature, C_T with C_T^I. Figure 3 depicts the complete CFU.

$$C_T' = C_T \times C_T^I \tag{10}$$

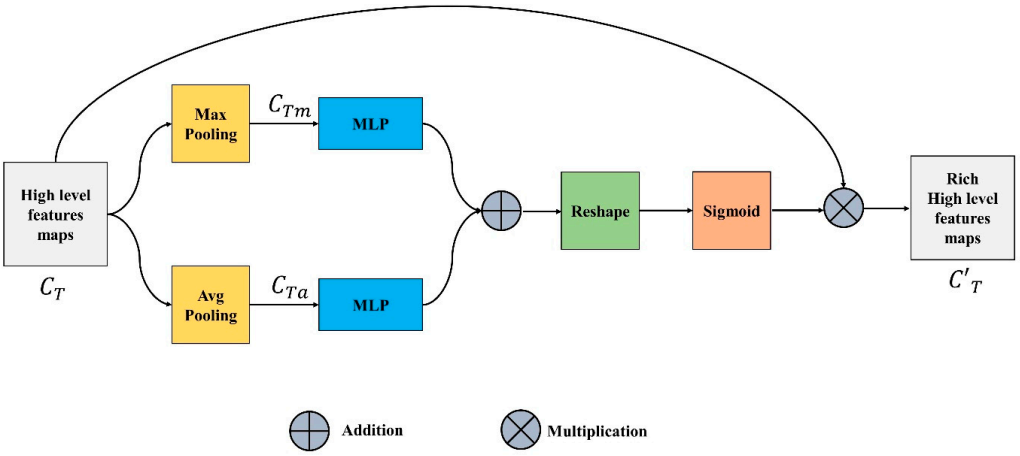


Figure 3. Channel Feature Unit.

3.4. Feature Integration Unit

There is a distinction in the level of feature representation between the two feature modules. Most of the spatial data acquired by the spatial path incorporate rich detailed information, while the output feature of the channel path includes primarily context information; therefore, directly combining or concatenating low-level and high-level features would inevitably lead to the degradation of information.

Because of this, we integrate low-level and high-level features using the feature integration unit (FIU), which properly merges various features comprehensively. To extract the most important pixel information from high-level features, global average pooling is deployed. Next, convolution, ReLU, and sigmoid functions are applied to F_H to acquire the feature indicator F_H shown in Equation (11). This F_H acts as a guide for low-level features.

$$F_H = \varsigma\left[conv\left\{\mathrm{P}_{ga}(C_T')\right\}\right] \tag{11}$$

Meanwhile, convolution operation and batch normalization are applied on low-level features to adjust and normalize the scale of channels and F_L is generated, represented in Equation (12).

$$F_L = conv[BN(C'_T)] \tag{12}$$

F_H is multiplied with F_L and finally is added to the result of multiplication to obtain the segmented output O_s as demonstrated in Equation (13).

$$O_s = C'_T + (F_H \times F_L) \tag{13}$$

Figure 4 demonstrate the complete FIU process.

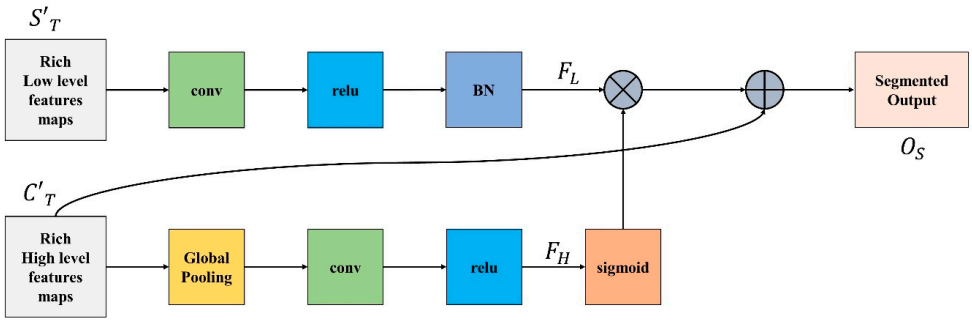


Figure 4. Feature Integration Unit.

3.5. Loss Function

In DL-based segmentation techniques, the loss function plays a critical role. A loss function is used to calculate how well the current output matches the target during training. Various segmentation tasks have proposed different loss functions. Dice is a loss function belonging to the class of region-based loss functions, which seek to reduce the mismatch or maximize the overlap areas between the ground-truth G and the anticipated segmentation S. In this study, the dice function is used as a loss function and is mentioned in Equation (14).

$$Loss = 1 - \frac{2\sum_{k=1}^{N}\sum_{l=1}^{M} g_l^k s_l^k}{\sum_{k=1}^{N}\sum_{l=1}^{M} g_l^k + \sum_{k=1}^{N}\sum_{l=1}^{M} s_l^k} \tag{14}$$

4. Experiments

In this section, the dataset used for both model training and testing is presented in detail. In addition, the hardware, hyperparameters, and evaluation metrics used to analyze the performance of the DL models are described.

4.1. Dataset

Data for this study was acquired from the National Heart Institute in Kuala Lumpur, Malaysia. It consists of 6000 two-dimensional images of apical four-chamber echocardiography images. These images were taken under protocol number RD5/04/15, which was authorized by Kuala Lumpur, Malaysia's National Heart Institute Research Ethics Board. The 2-dimensional echocardiogram was performed using a Philips IE33 ultrasound system equipped with an S5-1 (1.0–3.0 MHz) transducer. Every image had a dimension of 800 by 600 pixels, a resolution of 0.3 mm by 0.3 mm, with a frame rate ranging from 30 to 100 Hz. The irrelevant background was cropped out of all images, and the size was reduced to 512 × 512. The dataset is divided into three sections: 70% is used for training, 15% is used for testing, and the remaining 15% is used for validation. The binary masks are generated from the labelled LV. The labelled images were verified by medical professionals.

4.2. Network Training

A workstation with a Dell Core i7 Xeon E5-2620 CPU and an 11 GB Nvidia GeForce GTX 1080Ti GPU is used to train the network. An optimizer based on stochastic gradient descent with 0.9 momentum is chosen to be used during the training process. The initial learning rate is set to 10-4, the weight decay rate is set to 0.001, and the L2 regularization value is set to 0.0001, respectively. Each model is trained for 50 epochs, with a batch size of 32 and a shuffle after each epoch. The best-validated model configuration is chosen based on an examination of the training results using different sets of hyperparameters.

4.3. Evaluation Metrics

The trained network is assessed based on the 15% of test images reserved for testing. The trained model generates segmented binary masks for the test images, which are then compared to the ground truth binary mask for the test images. The proposed model's performance is evaluated and compared to that of other segmentation models using the most commonly used evaluation matrices for semantic segmentation, including the dice similarity coefficient (DSC), intersection over union (IoU), accuracy, recall, precision, and specificity [41,42].

5. Results

This section presents the findings and experimental outcomes of the LV segmentation. The proposed model's performance is compared with that of well-known DL segmentation models, Mask R-CNN, FCN, and DeepLab. The effectiveness of these models is investigated using the 15% test images.

5.1. Numerical Results

Table 1 displays a comparison of the segmentation performance of the proposed model using the five segmentation evaluation metrics. The table displays the DSC, IoU, accuracy, recall, and specificity average values for all test images. On test images, the proposed model obtains a DSC of 0.0.8637, an IoU of 0.9999, an accuracy of 0.9740, a recall value of 0.9886, and a specificity of 0.8352. Although DeepLab takes less time than other models, as was already determined, its overall performance is the lowest of all models. Due to its two-stage design, the Mask RCNN outperformed the FCN and DeepLab; however, among all segmentation architectures, the proposed model with a single stage had the best results.

Table 1. Mean and standard deviation values of evaluation metrics.

Model	DSC	IoU	Accuracy	Recall	Specificity	Precision
Proposed	0.9446 ±0.0319	0.8845 ±0.0536	0.9742 ±0.0114	0.9889 ±0.0081	0.8357 ±0.0817	0.9828 ±0.0088
Mask RCNN	0.9239 ±0.0232	0.8651 ±0.0457	0.9612 ±0.0116	0.9760 ±0.0084	0.8252 ±0.0845	0.9806 ±0.0087
FCN	0.8758 ±0.0439	0.7978 ±0.0536	0.9560 ±0.0115	0.9617 ±0.0178	0.8468 ±0.0587	0.9818 ±0.0095
DeepLab	0.8485 ±0.0409	0.7448 ±0.0620	0.9513 ±0.0092	0.9593 ±0.0167	0.7403 ±0.0787	0.9728 ±0.0098

Similarly, time spent on both training and segmentation of the LV from test images by the trained model is quantified for each model to evaluate the quickness of the models. Table 2 represents the timing information of the proposed model and other well-known segmentation models. From the data in, Table 2 it is clear that the proposed model not only required less time to train but was also fast enough to successfully segment the LV from the test images. The proposed model took 24.86 s to segment all 900 images or about 0.028 s per image. When compared to Mask RCNN and FCN, the proposed model's training and

testing times are significantly shorter. In contrast, DeepLab's training and testing times are shorter, but its performance is not competitive with the benchmark.

Table 2. Training and testing time of the models.

Model	Training Time	Segmentation Time of Test Images
Proposed	9 h. 53 min	24.86 s
Mask RCNN	12 h. 37 min	41.27 s
FCN	10 h. 57 min	31.97 s
DeepLab	6 h. 42 min	20.82 s

5.2. Visual Results

Figure 5 presents the visual results of the segmented output of all models. Original echocardiography images are in the leftmost column, and the ground truth labels are next to them. The segmented outputs of the proposed model, Mask RCNN, FCN, and DeepLab, are also depicted in column 3, column 4, column 5, and column 6 of Figure 5 respectively. DeepLab segmentation accuracy is not good enough, especially since the borderlines are inaccurate. The Mask RCNN and FCN border lines are quite precise; however, the shape of LV is not as accurate as the proposed model. The proposed model provided a highly accurate representation of the LV's form as well as its correct contour.

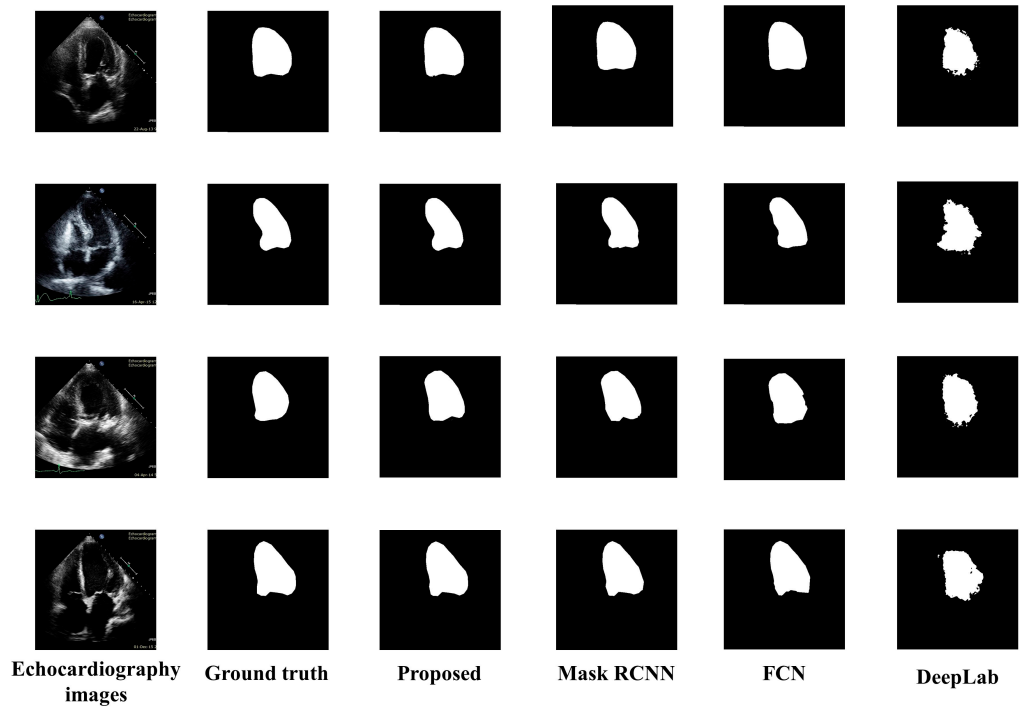


Figure 5. Visual comparisons of the proposed model to various segmentation models.

6. Discussion

In this work, a lightweight segmentation model is suggested, and its performance, as well as its training and testing times, are compared to DeepLab, FCN, and Mask RCNN. As can be seen in, Table 1 the suggested model achieved the highest levels of performance in terms of DSC, IoU, accuracy, recall, precision, and specificity. The proposed approach extracted features using the RestNet 50 model and then individually processed spatial and channel characteristics to maximize the usage of extracted features and information. On the other hand, the suggested model required much less time to train and segment test images compared to both FCN and Mask RCNN. This is because the spatial and channel

information is processed in parallel by the proposed model once the features are extracted from the backbone model.

In addition to this, the statistical analysis of the evaluation metrics is also presented. A boxplot was used to analyze and compare the distribution of evaluation metric values for all test images. Figure 6 depicts a boxplot of DSC values for test images. The proposed model outperforms competing segmentation methods in terms of the minimum and maximum value, the median, and the upper and lower quartiles. Similarly, the suggested model exhibits fewer outliers than the other models, indicating that its performance is consistent over the entire test dataset.

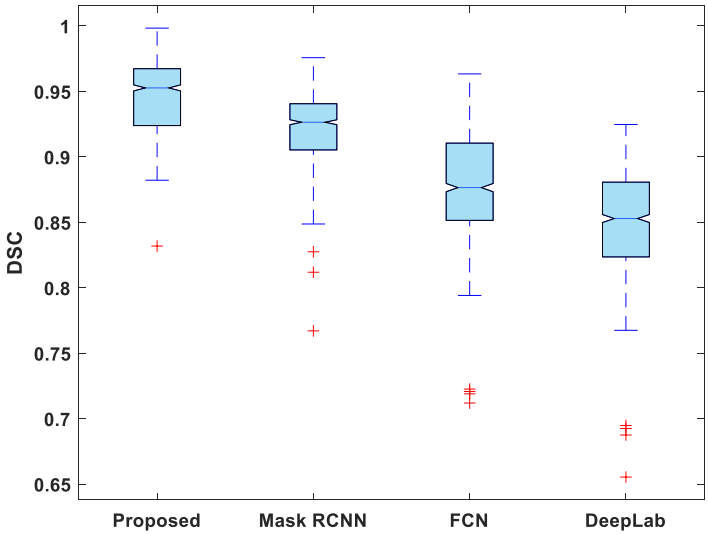


Figure 6. Boxplot of DSC.

Figures 7–11 display boxplots for IoU, accuracy, recall, precision, and specificity, respectively, from which it can be inferred that the behavior of the proposed model for these metrics is fairly similar to that of DSC. Furthermore, the proposed model's higher performance is also demonstrated by the lower skewness of the boxplot compared to the skewness of boxplots for the other models.

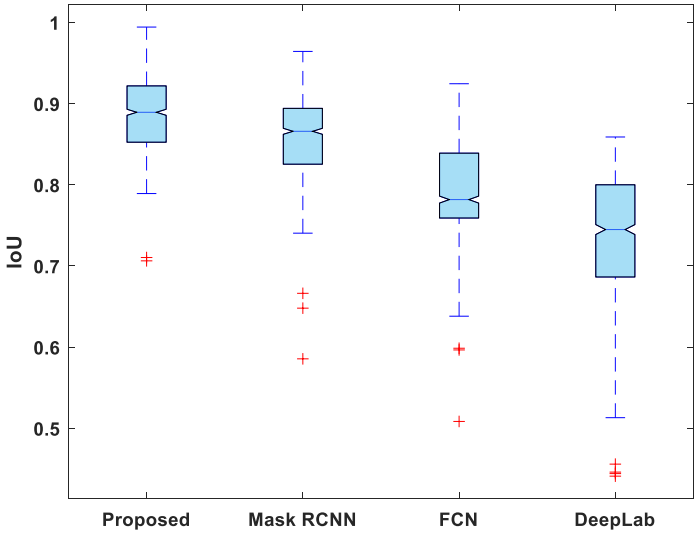


Figure 7. Boxplot of IoU.

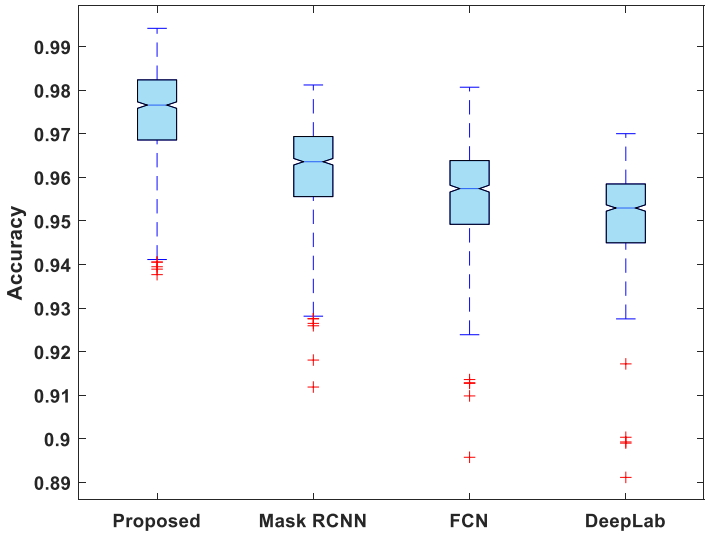


Figure 8. Boxplot of accuracy.

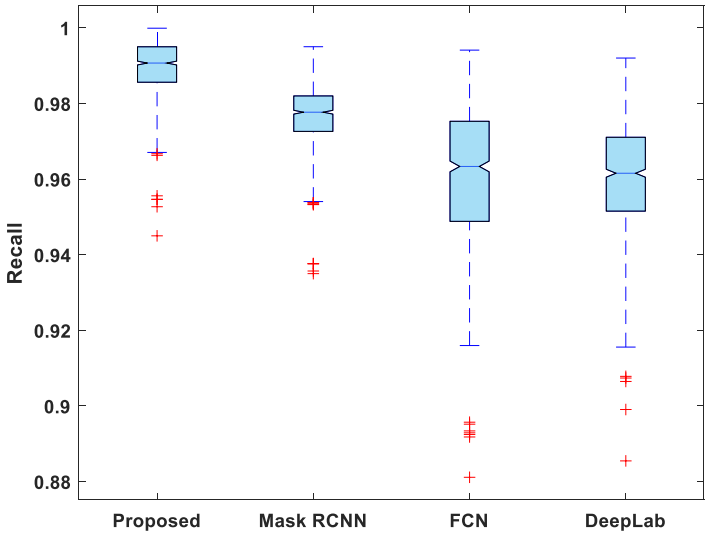


Figure 9. Boxplot of recall.

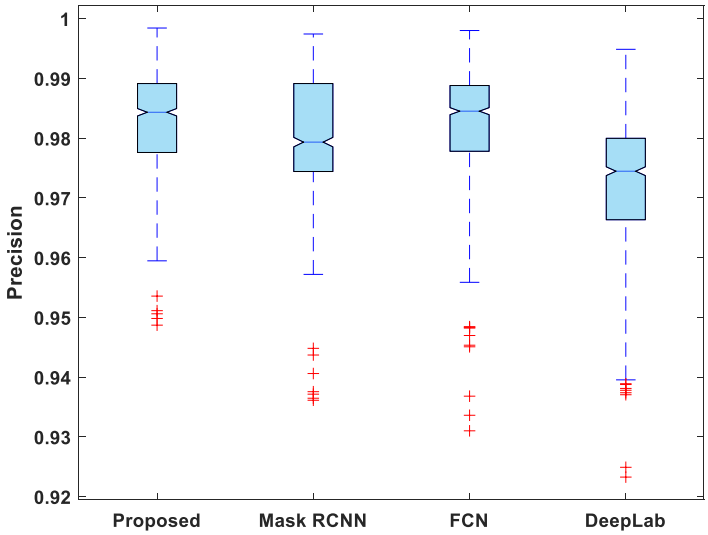


Figure 10. Boxplot of precision.

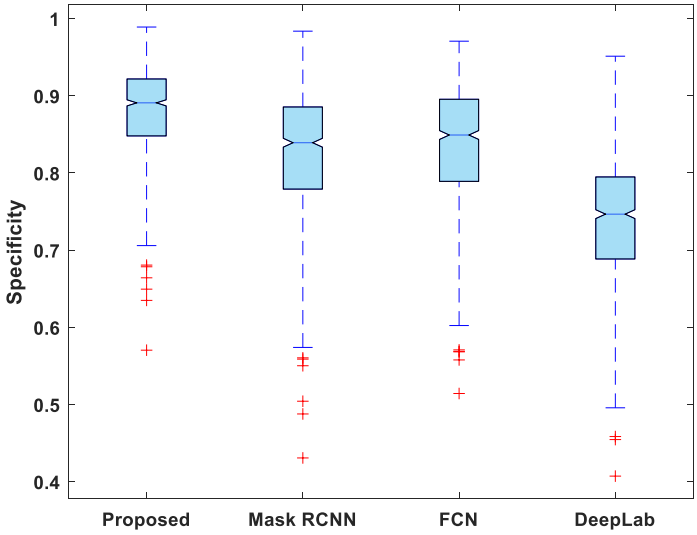


Figure 11. Boxplot of specificity.

7. Conclusions

A lightweight segmentation network was designed and applied in this study for automatic LV segmentation from echocardiography images. Utilizing restnet50 as the backbone model, a feature pyramid network is used to extract low and high-level features. The SFU uses pooling, convolution, batch normalization, and sigmoid processes to improve the spatial features generated by the backbone model. Similarly, the channel information is refined using global pooling and multi-layer perceptron in the channel feature unit. Using the feature integration unit, these two different level features are combined. The suggested model performed well not only in terms of segmentation accuracy but also in terms of speed. The segmenting of the test images and training of the model took less time. When compared to the well-known segmentation models Mask RCNN, FCN, and DeepLab, the proposed model scored the highest evaluation metrics with a DSC of 0.9446, IoU of 0.8445, an accuracy of 0.9742, recall of 0.9889, precision of 0.9828, and specificity of 0.8357.

Author Contributions: Conceptualization, M.A.S.; Data curation, A.K.; Formal analysis, M.A.S.; Investigation, M.A.S. and R.A.; Methodology, M.A.S.; Project administration, K.W.L.; Resources, A.K.; Software, M.A.S.; Supervision, K.W.L. and J.H.C.; Validation, J.H.C. and R.A.; Visualization, S.D. and Y.C.H.; Writing—original draft, M.A.S.; Writing—review and editing, K.W.L. and J.H.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Fundamental Research Grant Scheme (FRGS), Ministry of Higher Education Malaysia, and University Malaya under the Project FRGS/1/2019/TK04/UM/01/2.

Data Availability Statement: All the data are available upon request from the corresponding author.

Acknowledgments: This work is supported by the Fundamental Research Grant Scheme (FRGS), Ministry of Higher Education Malaysia, and University Malaya under the Project FRGS/1/2019/TK04/UM/01/2.

Conflicts of Interest: The authors declare that there are no conflict of interest regarding the publication of this study.

Abbreviations

Cardiovascular diseases (CVDs), left ventricle (LV), Deep learning (DL), Fully convolutional networks (FCN), convolutional neural networks (CNNs), spatial feature unit (SFU), channel features unit (CFU), feature integration unit (FIU), dice similarity coefficient (DSC), intersection over union (IoU).

References

1. Ali, M.K.; Bullard, K.M.; Saydah, S.; Imperatore, G.; Gregg, E.W. Cardiovascular and renal burdens of prediabetes in the USA: Analysis of data from serial cross-sectional surveys, 1988–2014. *Lancet Diabetes Endocrinol.* **2018**, *6*, 392–403. [CrossRef] [PubMed]
2. Townsend, N.; Wilson, L.; Bhatnagar, P.; Wickramasinghe, K.; Rayner, M.; Nichols, M. Cardiovascular disease in Europe: Epidemiological update 2016. *Eur. Heart J.* **2016**, *37*, 3232–3245. [CrossRef] [PubMed]
3. Khalil, A.; Ng, S.C.; Liew, Y.M.; Lai, K.W. An Overview on Image Registration Techniques for Cardiac Diagnosis and Treatment. *Cardiol. Res. Pract.* **2018**, *2018*, 1437125. [CrossRef] [PubMed]
4. Sugishita, Y.; Lambert, C.R. Textbook of clinical echocardiography. *Clin. Cardiol.* **1996**, *19*, 79. [CrossRef]
5. Ribes, R.; Kuschnir, K.; Luna, A.; Vilanova, J.C.; Jimenez-Hoyuela, J.M. (Eds.) *Learning Cardiac Imaging (Learning Imaging)*; Springer: Berlin, Heidelberg, Germany, 2010; ISBN 978-3540790822.
6. Hanif, M.; Nizar, A.; Khalil, A.; Chan, C.K.; Utama, N.P.; Lai, K.W. Pilot Study on Machine Learning for Aortic Valve Detection in Echocardiography Images. *J. Med. Imaging Health Inform.* **2018**, *8*, 1–6. [CrossRef]
7. Jiao, L.; Zhao, J. A Survey on the New Generation of Deep Learning in Image Processing. *IEEE Access* **2019**, *7*, 172231–172263. [CrossRef]
8. Kulathilake, K.A.S.H.; Abdullah, N.A.; Sabri, A.Q.M.; Bandara, A.M.R.R.; Lai, K.W. A review on self-adaptation approaches and techniques in medical image denoising algorithms. *Multimed. Tools Appl.* **2022**, *81*, 37591–37626. [CrossRef]
9. Serena Low, W.C.; Chuah, J.H.; Tee, C.A.T.H.; Anis, S.; Shoaib, M.A.; Faisal, A.; Khalil, A.; Lai, K.W. An Overview of Deep Learning Techniques on Chest X-Ray and CT Scan Identification of COVID-19. *Comput. Math. Methods Med.* **2021**, *2021*, 5528144. [CrossRef]
10. Suzuki, K. Overview of deep learning in medical imaging. *Radiol. Phys. Technol.* **2017**, *10*, 257–273. [CrossRef]
11. Anis, S.; Lai, K.W.; Chuah, J.H.; Ali, S.M.; Mohafez, H.; Hadizadeh, M.; Yan, D.; Ong, Z.-C. An Overview of Deep Learning Approaches in Chest Radiograph. *IEEE Access* **2020**, *8*, 182347–182354. [CrossRef]
12. Tajbakhsh, N.; Jeyaseelan, L.; Li, Q.; Chiang, J.N.; Wu, Z.; Ding, X. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Med. Image Anal.* **2020**, *63*, 101693. [CrossRef] [PubMed]
13. Liu, X.; Song, L.; Liu, S.; Zhang, Y. A review of deep-learning-based medical image segmentation methods. *Sustainability* **2021**, *13*, 1224. [CrossRef]
14. Lai, K.W.; Shoaib, M.A.; Chuah, J.H.; Ahmad Nizar, M.H.; Anis, S.; Woan Ching, S.L. Aortic Valve Segmentation using Deep Learning. In *Proceedings of the 2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), Langkawi Island, Malaysia, 1–3 March 2021*; IEEE: Piscataway, NJ, USA, 2021; pp. 528–532.
15. Dziri, H.; Cherni, M.A.; Ben-Sellem, D. New Hybrid Method for Left Ventricular Ejection Fraction Assessment from Radionuclide Ventriculography Images. *Curr. Med. Imaging Former. Curr. Med. Imaging Rev.* **2021**, *17*, 623–633. [CrossRef] [PubMed]
16. Jahanzad, Z.; Liew, Y.M.; Bilgen, M.; McLaughlin, R.A.; Leong, C.O.; Chee, K.H.; Aziz, Y.F.A.; Ung, N.M.; Lai, K.W.; Ng, S.C.; et al. Regional assessment of LV wall in infarcted heart using tagged MRI and cardiac modelling. *Phys. Med. Biol.* **2015**, *60*, 4015–4031. [CrossRef]
17. de Alexandria, A.R.; Cortez, P.C.; Bessa, J.A.; da Silva Félix, J.H.; de Abreu, J.S.; de Albuquerque, V.H.C. pSnakes: A new radial active contour model and its application in the segmentation of the left ventricle from echocardiographic images. *Comput. Methods Programs Biomed.* **2014**, *116*, 260–273. [CrossRef]
18. Renard, F.; Guedria, S.; Palma, N.D.; Vuillerme, N. Variability and reproducibility in deep learning for medical image segmentation. *Sci. Rep.* **2020**, *10*, 13724. [CrossRef]
19. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [CrossRef]
20. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin, Germany, 2015; Volume 9351, pp. 234–241. ISBN 9783319245737.
21. Koo, H.J.; Lee, J.-G.; Ko, J.Y.; Lee, G.; Kang, J.-W.; Kim, Y.-H.; Yang, D.H. Automated Segmentation of Left Ventricular Myocardium on Cardiac Computed Tomography Using Deep Learning. *Korean J. Radiol.* **2020**, *21*, 660. [CrossRef]
22. Ma, Z.; Wu, X.; Wang, X.; Song, Q.; Yin, Y.; Cao, K.; Wang, Y.; Zhou, J. An iterative multi-path fully convolutional neural network for automatic cardiac segmentation in cine MR images. *Med. Phys.* **2019**, *46*, 5652–5665. [CrossRef]
23. Zabihollahy, F.; Rajchl, M.; White, J.A.; Ukwatta, E. Fully automated segmentation of left ventricular scar from 3D late gadolinium enhancement magnetic resonance imaging using a cascaded multi-planar U-Net (CMPU-Net). *Med. Phys.* **2020**, *47*, 1645–1655. [CrossRef]
24. Yang, X.; Su, Y.; Tjio, G.; Yang, F.; Ding, J.; Kumar, S.; Leng, S.; Zhao, X.; Tan, R.-S.; Zhong, L. A Multi-channel Deep Learning Approach for Segmentation of the Left Ventricular Endocardium from Cardiac Images. In *Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019*; IEEE: Piscataway, NJ, USA, 2019; pp. 4016–4019.
25. Oktay, O.; Ferrante, E.; Kamnitsas, K.; Heinrich, M.; Bai, W.; Caballero, J.; Cook, S.A.; De Marvao, A.; Dawes, T.; O'Regan, D.P.; et al. Anatomically Constrained Neural Networks (ACNNs): Application to Cardiac Image Enhancement and Segmentation. *IEEE Trans. Med. Imaging* **2018**, *37*, 384–395. [CrossRef] [PubMed]

26. Leclerc, S.; Smistad, E.; Grenier, T.; Lartizien, C.; Ostvik, A.; Espinosa, F.; Jodoin, P.M.; Lovstakken, L.; Bernard, O. Deep Learning Applied to Multi-Structure Segmentation in 2D Echocardiography: A Preliminary Investigation of the Required Database Size. In *Proceedings of the 2018 IEEE International Ultrasonics Symposium (IUS), Kobe, Japan, 22–25 October 2018*; IEEE: Piscataway, NJ, USA, 2018; pp. 1–4. [CrossRef]
27. Leclerc, S.; Smistad, E.; Pedrosa, J.; Ostvik, A.; Cervenansky, F.; Espinosa, F.; Espeland, T.; Berg, E.A.R.; Jodoin, P.-M.; Grenier, T.; et al. Deep Learning for Segmentation using an Open Large-Scale Dataset in 2D Echocardiography. *IEEE Trans. Med. Imaging* **2019**, *38*, 2198–2210. [CrossRef]
28. Pesce, E.; Joseph Withey, S.; Ypsilantis, P.P.; Bakewell, R.; Goh, V.; Montana, G. Learning to detect chest radiographs containing pulmonary lesions using visual attention networks. *Med. Image Anal.* **2019**, *53*, 26–38. [CrossRef]
29. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention U-Net: Learning Where to Look for the Pancreas. *arXiv* **2018**, arXiv:1804.03999, arXiv:1804.03999. [CrossRef]
30. Vigneault, D.M.; Xie, W.; Ho, C.Y.; Bluemke, D.A.; Noble, J.A. Ω-Net (Omega-Net): Fully automatic, multi-view cardiac MR detection, orientation, and segmentation with deep neural networks. *Med. Image Anal.* **2018**, *48*, 95–106. [CrossRef] [PubMed]
31. Schlemper, J.; Oktay, O.; Schaap, M.; Heinrich, M.; Kainz, B.; Glocker, B.; Rueckert, D. Attention gated networks: Learning to leverage salient regions in medical images. *Med. Image Anal.* **2019**, *53*, 197–207. [CrossRef] [PubMed]
32. Cui, H.; Yuwen, C.; Jiang, L.; Xia, Y.; Zhang, Y. Multiscale attention guided U-Net architecture for cardiac segmentation in short-axis MRI images. *Comput. Methods Programs Biomed.* **2021**, *206*, 106142. [CrossRef]
33. Awasthi, N.; Vermeer, L.; Fixsen, L.S.; Lopata, R.G.P.; Pluim, J.P.W. LVNet: Lightweight Model for Left Ventricle Segmentation for Short Axis Views in Echocardiographic Imaging. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **2022**, *69*, 2115–2128. [CrossRef]
34. Hu, Y.; Xia, B.; Mao, M.; Jin, Z.; Du, J.; Guo, L.; Frangi, A.F.; Lei, B.; Wang, T. AIDAN: An Attention-Guided Dual-Path Network for Pediatric Echocardiography Segmentation. *IEEE Access* **2020**, *8*, 29176–29187. [CrossRef]
35. Shoaib, M.A.; Lai, K.W.; Khalil, A.; Chuah, J.H. Mask R-CNN for Segmentation of Left Ventricle. In *International Conference for Innovation in Biomedical Engineering and Life Sciences*; Springer: Cham, Switzerland, 2021; pp. 14–22.
36. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. In *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017*; IEEE: Piscataway, NJ, USA, 2017; pp. 2980–2988.
37. Shoaib, M.A.; Lai, K.W.; Chuah, J.H.; Hum, Y.C.; Ali, R.; Dhanalakshmi, S.; Wang, H.; Wu, X. Comparative studies of deep learning segmentation models for left ventricle segmentation. *Front. Public Health* **2022**, *10*, 2923. [CrossRef]
38. Ge, R.; Yang, G.; Chen, Y.; Luo, L.; Feng, C.; Zhang, H.; Li, S. PV-LVNet: Direct left ventricle multitype indices estimation from 2D echocardiograms of paired apical views with deep neural networks. *Med. Image Anal.* **2019**, *58*, 101554. [CrossRef]
39. Leclerc, S.; Jodoin, P.M.; Lovstakken, L.; Bernard, O.; Smistad, E.; Grenier, T.; Lartizien, C.; Ostvik, A.; Cervenansky, F.; Espinosa, F.; et al. RU-Net: A refining segmentation network for 2D echocardiography. In *Proceedings of the 2019 IEEE International Ultrasonics Symposium, Glasgow, UK, 6–9 October 2019*; IEEE: Piscataway, NJ, USA, 2019; pp. 1160–1163. [CrossRef]
40. Leclerc, S.; Smistad, E.; Østvik, A.; Cervenansky, F.; Espinosa, F.; Espeland, T.; Rye Berg, E.A.; Belhamissi, M.; Israilov, S.; Grenier, T.; et al. LU-Net: A Multistage Attention Network to Improve the Robustness of Segmentation of Left Ventricular Structures in 2-D Echocardiography. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **2020**, *67*, 2519–2530. [CrossRef] [PubMed]
41. Müller, D.; Soto-Rey, I.; Kramer, F. Towards a guideline for evaluation metrics in medical image segmentation. *BMC Res. Notes* **2022**, arXiv:2202.05273. [CrossRef] [PubMed]
42. Taha, A.A.; Hanbury, A. Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Med. Imaging* **2015**, *15*, 1–28. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.