Analysis of Right Ventricle segmentation in the End Diastolic and End Systolic cardiac phases using UNetbased models

Rania Mabrouk, Ramzi Mahmoudi, Asma Ammari, Rachida Saouli, Mohamed Hedi Bedoui

Analysis of Right Ventricle segmentation in the End Diastolic and End Systolic cardiac phases using UNet-based models.

Rania Mabrouk[1,2], Ramzi Mahmoudi[1,3], Asma Ammari[4], Rachida Saouli[4] and Mohamed Hedi Bedoui[1]

[1]Faculty of Medicine of Monastir, Medical Imaging Technology Lab - LTIM-LR12ES06, University of Monastir, Monastir, Tunisia.
{ranyamab96, hedi.bedoui2015}@gmail.com

[2]National engineering school of sfax, University of Sfax, Sfax, Tunisia.

[3]Gaspard-Monge computer-science laboratory, Paris-Est University, Mixed Unit CNRS-UMLV-ESIEE UMR8049, BP99, ESIEE Paris Cité Descartes, 93162 Noisy Le Grand, France.
ramzi.mahmoudi@esiee.fr

[4]Department of computer science, Laboratory LINFI, University of Biskra BP 145 RP 07000, Biskra, Algeria
asmaammari902@gmail.com ; rachida.saouli@esiee.fr

Abstract. Segmentation is an important task held to assess and analyze the heart's Right Ventricular (RV) function using CMR images. It has a major role in extracting important information which helps radiologists and doctors with the proper diagnosis. Several approaches have been proposed for RV segmentation showing great results in the End Diastolic (ED) phase but lower results in the End Systolic (ES) phase explained by the great variability of the complex shape of this chamber and its thin borders especially in the last phase. In this work, we aim to analyze the effect of short-axis slices from ED to ES phases on the segmentation task using a U-Net based architecture and two different datasets. Thus, a total of six models were trained to monitor the segmentation behavior.

Keywords: CMRI segmentation, right ventricle, end diastolic phase, end systolic phase, U-Net.

1 Introduction

The clinical importance of the Right Ventricle in cardiovascular diseases [1] has been encouraging to assess its function for a better and a more accurate diagnosis [2], [3]. Various imaging modalities are used for the RV evaluation where Cardiovascular Magnetic Resonance Imaging (CMRI) is the gold standard reference [4],[5]. To analyze the RV function, radiologists have to delineate its boundaries over the entire slices which is a time-consuming task. For this reason,

automatic segmentation of this cardiac cavity has been studied using multiple approaches [6]. Despite the inspiring results obtained in the End Diastolic (ED) phase, lower results were detected in the End Systolic (ES) phase for many proposed approaches [7]. Those results are explained by the great variability of the complex shape of this cavity and its thin borders especially in the ES phase where the chambers are found considerably narrowed.

This paper aims to analyze the impact of short-axis slices from ED to ES phases on the segmentation task. For this reason, we proposed a U-Net based architecture and used two different datasets (a private one and a public one). A total of six models were trained to observe the segmentation performance.

The remainder of this paper is organized as follows: In section 2, a brief literature overview is presented. The used datasets and the proposed architecture are detailed in section 3. The experimental results including a comparison with similar works are discussed in section 4. Finally, section 5 concludes the paper and proposes possible improvements.

2 Related Works

To tackle the challenges of Right Ventricle segmentation, various works were proposed employing different segmentation techniques [8]. As reviewed in [6] and [9], the most recently proposed methods are more oriented to use deep learning techniques. In fact, Good progress in the medical imaging field has been reached thanks to the introduction of Artificial Intelligence technologies that became a popular approach for detection and segmentation problems due to their powerful feature representation [10]. A Multi-Centre, Multi-Vendor and Multi-Disease Cardiac Segmentation Challenge was organized as part of the MICCAI 2020 Conference [7] where a total of fourteen teams submitted different techniques for CMR images segmentation, including Left Ventricle segmentation, myocardium segmentation and Right Ventricle segmentation, using the same proposed dataset. As we are only interested in the RV segmentation, we note that the best dice coefficient reached in the challenge for this task was 0.91 at the End Diastolic phase and 0.86 at the End Systolic phase.

3 Proposed Method

3.1 Datasets Description

We adopted two different datasets to monitor the behavior of U-Net-based models in accordance with data.

The first one is “LabTIM-RV” private dataset proposed within our laboratory collected from the University Hospital of Fattouma Bourguiba (Monastir, Tunisia)

in collaboration with its radiology service. It constitutes a total of 160 patients with a total number of 3528 labeled RV CMR images at both ED and ES phases. We subdivided this dataset into two other subsets each containing a total of 1659 RV labeled images at ED and ES phases separately.

The second one is the publicly available dataset used in the Multi-Centre, Multi-Vendor and Multi-Disease Cardiac Segmentation M&Ms Challenge which was organized as part of the MICCAI 2020 Conference [7]. It contains a total of 160 patients with a total number of 3554 short-axis CMR images along with their corresponding ground truth images of Left Ventricle LV, LV myocardium, and Right Ventricle RV. As we are interested in segmenting the Right Ventricle, we only extracted its labels by applying a simple threshold on the original ground truth images. We subdivided this dataset as well into two other subsets each containing a total of 1777 RV labeled images at ED and ES phases separately.

Both datasets were resized to 256x256 as they had a wide variety of dimensions ranging from 174x208 to 512x512.

3.2 Architecture

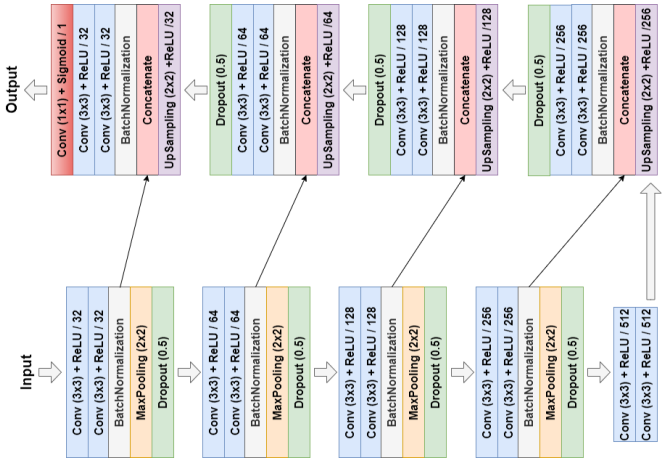

Fig. 1. The proposed U-Net-based architecture.

The U-Net-based architecture implemented for the analysis in this work consists of five blocks encoding path and a symmetric five blocks decoding path as detailed in Fig.1. At each level of the encoder, a convolution operation and a ReLU activation function were applied two times consecutively followed by a batch normalization operation, a max-pooling operation then a dropout layer before moving to the next level. The decoder, then, recovers the original input size by applying the same sequence of operations with replacing the max-pooling operation with the transposed convolution as an upsampling operation at every level. The corresponding feature from the encoder is concatenated to the decoder's block input as well. A 1x1 convolution, with a sigmoid activation function, was

then added at last for the generation of the final binary prediction map. The convolutions were applied with a kernel size of 3x3 and the transposed convolutions were applied with a kernel size of 2x2 in a stride of 2x2.
A total of six models were trained using this architecture with different datasets.
The first model (model1) was trained using the entire private dataset described above. The second model (model2) was trained using a subset that contains only the End Diastolic slices of the private dataset whereas the third model (model3) was trained using another subset that contains only the End Systolic slices of the private dataset. The forth model (model4) was trained using the entire public dataset described above as well. The fifth model (model5) was trained using a subset that contains only the End Diastolic slices of the public dataset whereas the sixth model (model6) was trained using another subset that contains only the End Systolic slices of the public dataset.
Each used dataset was subdivided into 10 patients for the testing process and 150 patients for the training process that itself was partitioned into 70 % for the training set and 30 % for the validation set as detailed in Table1.

Table 1. Details of the different used datasets in the learning and test processes.

	Dataset	Total imgs	Train	Validation	Test
Private dataset	Entire Set	3528	2322	996	210
	ED subset	1764	1161	498	105
	ES subset	1764	1161	498	105
Public dataset	Entire Set	3554	2336	1002	216
	ED subset	1777	1168	501	108
	ES subset	1777	1168	501	108

All networks were trained using the same hyper parameters including Dice Loss as a loss function, 32 batch size ,100 epochs and Adam optimizer with 0.0005 as a learning rate. To further optimize the training procedure, we used the cosine annealing scheduler, implemented as a custom callback, where the learning rate ranges between 0.0005 and 0.0001.

4 Experimental Results and Discussion

In this section, we provide a detailed experimental analysis of the proposed models that demonstrates quantitative results held with each of the datasets used in the training phase. We compare at last our results with state-of-the-art methods.

4.1 Evaluation Metrics

Various metrics are being used by the research community for medical image analysis to quantify the performance of segmentation models among which we can cite Dice Coefficient, Intersection over Union, Precision and Recall as the most

popular choice.
Four measures are required to calculate these metrics:
True Positive (TP): is the number of RV pixels being correctly identified as RV pixels.
True Negative (TN): is the number of non-RV pixels being correctly identified as non-RV pixels.
False Positive (FP): is the number of non-RV pixels being wrongly identified as RV pixels.
False Negative (FN): is the number of RV pixels being wrongly identified as non-RV pixels.

Dice Coefficient: is the overlap ratio between the prediction and the ground truth with giving more weight to the intersection between them two and defined in (1). It ranges between 0 and 1 and the higher the value is, the better the segmentation result.

$$Dice\ Coefficient = 2TP / (2TP + FP + FN) \quad (1)$$

Intersection over Union (IoU): it measures the overlap between the prediction and the ground truth and is defined in (2).

$$Dice\ Coefficient = TP / (TP + FP + FN) \quad (2)$$

Precision: is a measure of exactness calculated as the ratio of true positive predictions divided by the number of predicted positives and defined in (3).

$$Precision = TP / (TP + FP) \quad (3)$$

Recall: is a measure of completeness calculated as the ratio of true positive predictions divided by the number of actual positives and defined in (4).

$$Recall = TP / (TP + FN) \quad (4)$$

4.2 Quantitative evaluation

In this section a comparative experiment is presented to address RV segmentation challenging issues caused by the shape variation from End Diastolic and End Systolic slices. Consequently, we decided to study the impact of learning each slice-level separately. For that, a total of six models were trained using the different datasets and the same U-Net based architecture as detailed above.
To evaluate and study the behavior of the RV segmentation among the different datasets, the dice coefficient and dice loss are computed.

Private Dataset

Fig2. demonstrates the dice and loss curves of training and validation of the RV segmentation models, where the whole private dataset was used for training first (a) then the ED (b) and ES (c) private subsets were considered next separately.

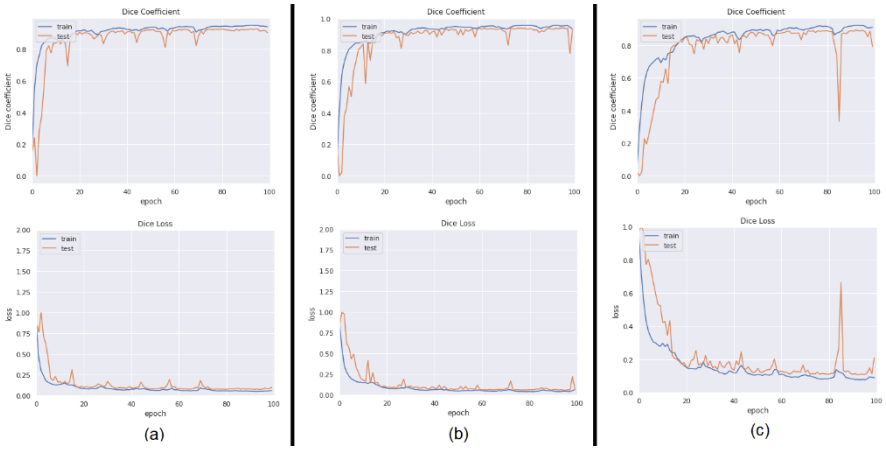

Fig. 2. Dice Coefficient and Dice Loss Training curves of the RV segmentation models of the private dataset. (a) model1: using the entire dataset. (b) model2: using the ED subset. (c) model3: using the ES subset.

The third model presented by Fig.2(c) seems to be confused more than the other models. Whereas the first model presented by Fig.2(a) demonstrates a better behavior along the training process where the Dice Coefficient curves/ Dice Loss curves continued to increase/ decrease to a point of stability.

Public Dataset

Fig3. demonstrates the dice and loss curves of training and validation of the RV segmentation models, where the whole public dataset was used for training first (a) then the ED (b) and ES (c) public subsets were considered next separately.

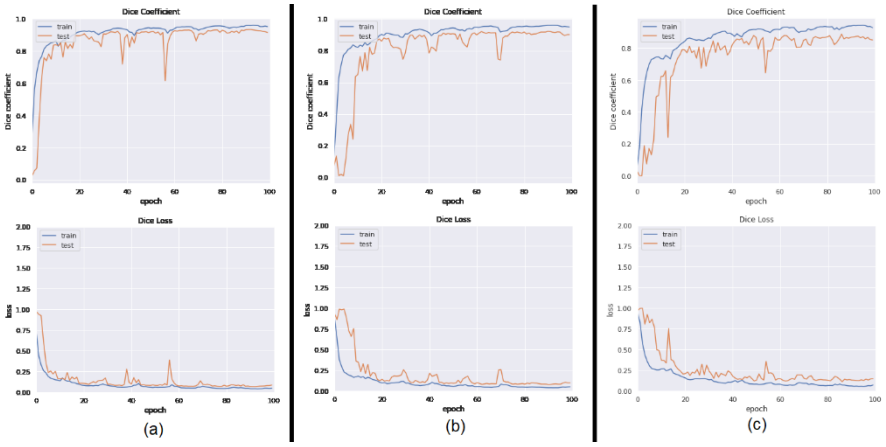

Fig. 3. Dice Coefficient and Dice Loss Training curves of the RV segmentation models of the public dataset. (a) using the entire dataset. (b) using the ED phase dataset. (c) using the ES phase dataset.

Same to the previous experiments, the third model presented by Fig.3(c) seems to

be confused more than the other models. Whereas the first model presented by Fig.3(a) demonstrates a better behavior along the training process where the Dice Coefficient curves/ Dice Loss curves continued to increase/ decrease to a point of stability. Therefore, it is safe to assume that the challenging issues related to the End systolic phase may influence the segmentation results and even affect the segmentation of other slices.

Table 2. Comparison of RV segmentation performance of the different models in the validation phase, in terms of Dice Coefficient, IoU, Precision and Recall.

	Metric	Dice	IoU	Precision	Recall
Private dataset	Model1	0.9272	0.8653	0.9378	0.9204
	Model2	0.9437	0.8940	0.9472	0.9445
	Model3	0.8931	0.8079	0.8944	0.8968
Public dataset	Model4	0.9322	0.8745	0.9362	0.9326
	Model5	0.9221	0.8569	0.9374	0.9116
	Model6	0.8514	0.7447	0.9463	0.7803

Table 3. Comparison of RV segmentation performance of the different models in the test phase, in terms of Dice Coefficient, IoU, Precision and Recall.

	Metric	Dice	IoU	Precision	Recall
Private test set	Model1	0.9058	0.8325	0.9603	0.8661
	Model2	0.9241	0.8609	0.9872	0.8727
	Model3	0.8243	0.7061	0.8987	0.7683
Public test set	Model4	0.8890	0.8125	0.9603	0.8661
	Model5	0.8704	0.8124	0.8855	0.8597
	Model6	0.7394	0.6103	0.9216	0.6346

Table2/Table3 reports a comparison of RV segmentation performance in the validation/test phase, in terms of Dice Coefficient, IoU, Precision and Recall when training with the whole private dataset (model1), with the private ED subset (model2), with the private ES subset (model4), with the entire public dataset (model4), with the public ED subset (model5) and with the public ES subset (model6).

Training with the private ED subset (model2) shows the best performance as it reached a validation Dice Coefficient of 0.9437 and test Dice Coefficient of 0.9241 which are the highest values in comparison with model1 and model3. Whereas training with the public whole dataset (model4) shows the best performance as it reached a validation Dice Coefficient of 0.9322 and test Dice Coefficient of 0.8890 which are the highest values in comparison with model5 and model6.

Training with the private ES subset (model3) and the public one (model6) both show the worst performance as model3 and model6 reached a validation Dice Coefficient of 0.8931 and 0.8514 respectively and test Dice Coefficient of 0.8243 and 0.7394 respectively which are the lowest values in comparison with the other models.

Table 4. Comparison of RV segmentation performance in the validation phase, in terms of Dice Coefficient, IoU, Precision and Recall at the end-systolic (ED) and the end-diastolic (ES) phases, when training with the whole private dataset (model1) and the entire public dataset (model4).

	Dice		IoU		Precision		Recall	
	ED	ES	ED	ES	ED	ES	ED	ES
Model1	0.9586	0.9148	0.9208	0.8442	0.9731	0.9181	0.9469	0.9159
Model4	0.9587	0.9352	0.9224	0.8821	0.9709	0.9338	0.9504	0.9424

Table 5. Comparison of RV segmentation performance in the test phase, in terms of Dice Coefficient, IoU, Precision and Recall at the end-systolic (ED) and the end-diastolic (ES) phases, when training with the whole private dataset (model1) and the entire public dataset (model4).

	Dice		IoU		Precision		Recall	
	ED	ES	ED	ES	ED	ES	ED	ES
Model1	0.9292	0.8783	0.8694	0.7867	0.9959	0.9179	0.8737	0.8470
Model4	0.8893	0.8416	0.8081	0.7453	0.9362	0.8904	0.9326	0.8098

Table4/Table5 reports a comparison of RV segmentation performance in the validation/test phase, in terms of Dice Coefficient, IoU, Precision and Recall at the end-systolic (ED) and the end-diastolic (ES) phases, when training with the whole private dataset (model1) and the entire public dataset (model4). Both tables demonstrate that the segmentation performance in the End Systolic phase is lower than the End Diastolic phase for all computed metrics.

Table 6. Comparison of End Systolic slices segmentation performance in the validation phase using the different models, in terms of Dice Coefficient, IoU, Precision and Recall.

	Metric	Dice	IoU	Precision	Recall
Private dataset	**Model1**	**0.9148**	**0.8442**	**0.9181**	**0.9159**
	Model2	0.8317	0.7200	0.7949	0.8875
	Model3	0.8931	0.8079	0.8944	0.8968
Public dataset	**Model4**	**0.9352**	**0.8821**	**0.9338**	**0.9424**
	Model5	0.8591	0.7601	0.8374	0.8948
	Model6	0.8514	0.7447	0.9463	0.7803

Table 7. Comparison of End Systolic slices segmentation performance in the test phase using the different models, in terms of Dice Coefficient, IoU, Precision and Recall.

	Metric	Dice	IoU	Precision	Recall
Private test set	**Model1**	**0.8783**	**0.7867**	**0.9179**	**0.8470**
	Model2	0.8517	0.7490	0.8555	0.8536
	Model3	0.8243	0.7061	0.8987	0.7683
Public test set	**Model4**	**0.8416**	**0.7453**	**0.8904**	**0.8098**
	Model5	0.8012	0.6739	0.8382	0.7768
	Model6	0.7394	0.6103	0.9216	0.6346

Table6/Table7 reports a comparison of End Systolic slices segmentation performance in the validation/test phase, in terms of Dice Coefficient, IoU, Precision

and Recall when training with the whole private dataset (model1), with the private ED subset (model2), with the private ES subset (model4), with the entire public dataset (model4), with the public ED subset (model5) and with the public ES subset (model6).
Training with the private whole dataset (model1) and the public one (model4) both show the best performance as model1 and model4 reached a validation Dice Coefficient of 0.9148 and 0.9352 respectively and test Dice Coefficient of 0.8783 and 0.8416 respectively which are the highest values in comparison with the other models.
These results don't show a better impact of excluding End Systolic slices on the segmentation performance.

4.3 Qualitative evaluation

The Qualitative evaluation proved that the predicted masks of the different six models give a good agreement with the reality as the results are notably close to the original ones as demonstrated in Figures below displaying each four sample images from the test set.

4.4 Comparison with state of the art methods

In this section, we quantitatively compare the performance of our proposed RV segmentation model, trained using the whole public dataset (model4), with fourteen state of the art methods submitted within the M&Ms Challenge organized as part of the MICCAI 2020 Conference [7].
The dice coefficients obtained within the challenge range between 0.910 and 0.552 in the End Diastolic phase which are lower than the dice coefficient we obtained using our model (0.9587). The dice coefficients obtained in the End Systolic phase range between 0.860 and 0.517 which are also lower than the dice coefficient we obtained using our model (0.9352).

5 Conclusion

Inspite of the promising results reached with our proposed U-Net architecture that surpassed state of the art methods, we conclude that U-Net alone is still insufficient to tackle the RV segmentation challenging issues in the ES phase. Hence, further improvements and other approaches are needed. It may be wiser to propose a particular approach for each phase separately. In addition, choosing a specific segmentation method is not the only concern but it would be interesting, as well, to study how to efficiently exploit the CMRI available slices to achieve higher results for the entire cardiac short-axis sequence. In fact, data augmentation, preprocessing and picking the most relevant clinical cases can have a great positive influence in the segmentation process.

References

1. Sheehan, F., Redington, A.: The right ventricle: anatomy, physiology and clinical imaging. Heart. 94, 1510–1515 (2008). https://doi.org/10.1136/hrt.2007.132779.
2. Goetschalckx, K., Rademakers, F., Bogaert, J.: Right ventricular function by MRI. Current Opinion in Cardiology. 25, 451–455 (2010). https://doi.org/10.1097/HCO.0b013e32833b78e6.
3. Tavano, A., Maurel, B., Gaubert, J.-Y., Varoquaux, A., Cassagneau, P., Vidal, V., Bartoli, J.-M., Moulin, G., Jacquier, A.: MR imaging of arrhythmogenic right ventricular dysplasia: What the radiologist needs to know. Diagnostic and Interventional Imaging. 96, 449–460 (2015). https://doi.org/10.1016/j.diii.2014.07.009.
4. Steen, H., Nasir, K., Flynn, E., El-Shehaby, I., Lai, S., Katus, H.A., Bluemcke, D., Lima, J. a. C.: Is magnetic resonance imaging the "reference standard" for cardiac functional assessment? Factors influencing measurement of left ventricular mass and volumes. Clin Res Cardiol. 96, 743–751 (2007). https://doi.org/10.1007/s00392-007-0556-2.
5. Caudron, J., Fares, J., Lefebvre, V., Vivier, P.-H., Petitjean, C., Dacher, J.-N.: Cardiac MRI assessment of right ventricular function in acquired heart disease: factors of variability. Acad Radiol. 19, 991–1002 (2012). https://doi.org/10.1016/j.acra.2012.03.022.
6. Ammari, A., Mahmoudi, R., Hmida, B., Saouli, R., Bedoui, M.H.: A review of approaches investigated for right ventricular segmentation using short-axis cardiac MRI. IET Image Processing. (2021). https://doi.org/10.1049/ipr2.12165.
7. Campello, V.M., Gkontra, P., Izquierdo, C., Martín-Isla, C., Sojoudi, A., Full, P.M., Maier-Hein, K., Zhang, Y., He, Z., Ma, J., Parreño, M., Albiol, A., Kong, F., Shadden, S.C., Acero, J.C., Sundaresan, V., Saber, M., Elattar, M., Li, H., Menze, B., Khader, F., Haarburger, C., Scannell, C.M., Veta, M., Carscadden, A., Punithakumar, K., Liu, X., Tsaftaris, S.A., Huang, X., Yang, X., Li, L., Zhuang, X., Viladés, D., Descalzo, M.L., Guala, A., Mura, L.L., Friedrich, M.G., Garg, R., Lebel, J., Henriques, F., Karakas, M., Çavuş, E., Petersen, S.E., Escalera, S., Seguí, S., Rodríguez-Palomares, J.F., Lekadir, K.: Multi-Centre, Multi-Vendor and Multi-Disease Cardiac Segmentation: The M amp;Ms Challenge. IEEE Transactions on Medical Imaging. 40, 3543–3554 (2021). https://doi.org/10.1109/TMI.2021.3090082.
8. Petitjean, C., Zuluaga, M.A., Bai, W., Dacher, J.-N., Grosgeorge, D., Caudron, J., Ruan, S., Ayed, I.B., Cardoso, M.J., Chen, H.-C., Jimenez-Carretero, D., Ledesma-Carbayo, M.J., Davatzikos, C., Doshi, J., Erus, G., Maier, O.M.O., Nambakhsh, C.M.S., Ou, Y., Ourselin, S., Peng, C.-W., Peters, N.S., Peters, T.M., Rajchl, M., Rueckert, D., Santos, A., Shi, W., Wang, C.-W., Wang, H., Yuan, J.: Right ventricle segmentation from cardiac MRI: A collation study. Medical Image Analysis. 19, 187–202 (2015). https://doi.org/10.1016/j.media.2014.10.004.
9. Chen, C., Qin, C., Qiu, H., Tarroni, G., Duan, J., Bai, W., Rueckert, D.: Deep Learning for Cardiac Image Segmentation: A Review. Frontiers in Cardiovascular Medicine. 7, (2020).
10. Voulodimos, A., Doulamis, N., Doulamis, A., Protopapadakis, E.: Deep Learning for Computer Vision: A Brief Review. Computational Intelligence and Neuroscience. 2018, e7068349 (2018). https://doi.org/10.1155/2018/7068349.