Pérez-Pelegrí, M.; Monmeneu, JV.; López-Lereu, MP.; Ruiz-España, S.; Del-Canto, I.; Bodi, V.; Moratal, D. (2020). PSPU-Net for Automatic Short Axis Cine MRI Segmentation of Left and Right Ventricles. IEEE Computer Society. 1048-1053. https://doi.org/10.1109/BIBE50027.2020.00177

PSPU-Net for automatic short axis cine MRI segmentation of left and right ventricles

Manuel Pérez-Pelegrí
Center for Biomaterials and Tissue Engineering
Universitat Politècnica de València
Valencia, Spain
mapepe18@gmail.com

José V. Monmeneu
Unidad de Imagen Cardíaca
ERESA
Valencia, Spain
jmonmeneu@eresa.com

María P. López-Lereu
Unidad de Imagen Cardíaca
ERESA
Valencia, Spain
plereu@eresa.com

Silvia Ruiz-España
Center for Biomaterials and Tissue Engineering
Universitat Politècnica de València
Valencia, Spain
silviaruiz.es@gmail.com

Irene Del-Canto
Center for Biomaterials and Tissue Engineering
Universitat Politècnica de València
Valencia, Spain
irdecan@upv.es

Vicente Bodí
Department of Medicine
Universitat de València
Valencia, Spain
vicentbodi@hotmail.com

David Moratal
Center for Biomaterials and Tissue Engineering
Universitat Politécnica de València
Valencia, Spain
dmoratal@eln.upv.es

***Abstract*— Characterization of the heart anatomy and function is mostly done with magnetic resonance image cine series. To achieve a correct characterization, the volume of the right and left ventricle need to be segmented, which is a time-consuming task. We propose a new convolutional neural network architecture that combines U-net with PSP modules (PSPU-net) for the segmentation of left and right ventricle cavities and left ventricle myocardium in the diastolic frame of short-axis cine MRI images and compare its results against a classic 3D U-net architecture. We used a dataset containing 399 cases in total. The results showed higher quality results in both segmentation and final volume estimation for a test set of 99 cases in the case of the PSPU-net, with global dice metrics of 0.910 and median absolute relative errors in volume estimations of 0.026 and 0.039 for the left ventricle cavity and myocardium and 0.051 for the right ventricles cavity.**

Keywords—MRI, U-net, PSP, segmentation, deep learning, left ventricle, right ventricle, volume estimation.

I. INTRODUCTION

Cardiovascular diseases are one of the most important public health problems in advanced countries and are the main cause of death [1]. As such, the correct characterization of the heart structure and function is vital for a correct assessment and diagnosis. In this setting, magnetic resonance imaging (MRI) is one the most used diagnostic tools for cardiac structure and function assessment, it being a high-resolution and non-invasive technique.

In the characterization of heart disease for diagnostic purposes, some of the main parameters are those related to volume estimation of the heart main regions: the left ventricle (LV) and the right ventricle (RV). In order to measure these parameters, it is necessary to correctly segment said regions and its conforming tissues, which is a tedious and time-consuming task.

In this context deep learning techniques have been used in the past years in order to help characterize the heart in MRI images, giving special attention to the LV segmentation [2,3,4]. In this work we propose a new architecture based on the U-net model [5] and with the addition of Parse scene parsing blocks (PSP) [6] and compare it with a basic 3D U-net for the segmentation of the left ventricle cavity, left ventricle myocardium and right ventricle cavity in the diastolic frame of MRI short axis cine-cardiac series.

II. METHOD

A. Image dataset

Our dataset consisted of 399 short-axis stacks of MRI covering the left and right ventricles. For the experiments we used only the imaging time frames corresponding to the systole and diastole. All patients gave written consent and the study was approved by the Medical Ethical Committee of our hospital (Hospital Clínico Universitario de Valencia, Valencia, Spain). Imaging was performed in breath-hold using a 1.5T MRI scanner (Sonata Magnetom, Siemens, Erlangen, Germany), flip angle: 58°, repetition time: 52.92 ms, echo time: 1.25 ms.

The in-plane resolution was variable among the cases, ranging from 0.57×0.57 mm^2 to 1.09×1.09 mm^2, with a slice thickness of 7 mm and a spacing between slices of 3 mm (in all cases). The resulting image sizes varied from 144 × 144 to 256 x 256 and the number of slices ranged from 8 to 14. All the images were resampled to a constant in-plane spatial resolution of 1 mm^2 with an image size of 176 × 176. Although some images had to be cropped and others had to be zero-padded, still in every case the region of interest remained with a great margin as all the images placed the heart in the center region. The z-axis was left untouched in the resampling process.

Every case was classified in one of 11 categories described in table 1. This was done to ensure that the training, validation and test sets had the same distribution respect to the diagnosis.

Finally, all the images were randomly split in training (260 cases, 65% of the cases), validation (40 cases, 10% of the cases) and test (99 cases, 25% of the cases) sets.

B. Tissue Labels

Each diastolic frame had been previously manually segmented by mutual consent of two expert cardiologists with

Conselleria d'Educació, Investigació, Cultura i Esport, and Agencia Valenciana de la Innovación, Generalitat Valenciana. Centro para el Desarrollo Tecnológico Industrial, Ministerio de Ciencia, Innovación y Universidades

TABLE 1. Categories by pathology in the dataset and the number of occurences

Categories	Number of cases
Normal cases, no pathology	48
Presence of necrosis	14
Presence of fibrosis	12
Presence of ischemia	10
Functional affection of LV (ejection fraction lower than normal and/or affected segmental contractility)	23
Functional affection of RV (ejection fraction lower than normal and/or affected segmental contractility)	2
Functional affection of LV and RV	137
Functional affection of LV and presence of fibrosis/necrosis/ischemia	45
Functional affection of RV and presence of fibrosis/necrosis/ischemia	4
Functional affection of RV and LV and presence of fibrosis/necrosis/ischemia	95
Other cases that do not fall in any other category	9

more than 12 years of experience. The segmentations originally only included the contours of each tissue, we processed the segmentations to generate full volumetric segmentations and associated values of 1 to LV cavity, values of 2 to LV myocardium and values of 3 to RV cavity. These volumetric segmentations were used as gold standard to train and test the networks. The neural networks were also tested against the systolic frame to see performance with images of different nature but similar features. In the case of the systolic frame the segmentations only included the LV and RV cavities (labels 1 and 2 respectively). In this case the segmentation was available in 98 of the test cases.

C. Convolutional neural network architecture

Two different U-net based architectures were designed and trained. The first one is a basic U-net for 3D segmentation (U-net 3D) that had been previously trained with the dataset. The U-net 3D is similar to the classical U-net for 3D segmentation [7], but in our design we established more downsampling and upsampling steps (a total of 4), and these steps worked only in the 2D plane, so the z-dimension remains the same size along all the network. Every convolution was followed by batch normalization [8] with a ReLu activation function. The final result of the network is a feature map of 4 channels, each representing the probability inferred for each label (3 tissue labels and the background). The total number of trainable parameters was of 87.51 million, and the neural network occupied 1.03 GB. Fig. 1 shows the architecture for the 3D U-net.

The PSPU-net includes some notable changes. Even though it takes the same inputs as the 3D U-net (176 × 176 × 3), in this architecture the convolutions are 3 × 3 × 1, which makes this architecture a 2D U-net in the sense that it only uses 2D information. Additionally, the number of filters is halved. The network still outputs the segmentation in 3D (176 × 176 × 3). The PSPU-net also includes PSP modules in the place of the skip connections.

The PSP-modules are designed as in Fig. 2. It shows the PSP-module used in the highest level (4 paths). The modules design changes depending on the number of paths employed, eliminating the path with the highest sampling rate and doubling the number of filters with respect the previous PSP module with a higher path number. One can easily view this design with the example of the 3-path module: parting from the 4-path module, the x16 path is suppressed and the number of filters is doubled.

The classic U-net can only analyze the image at different scales following the contracting path, the skip connections allows for it to recover original spatial information, but this information is never furthered analyzed at different scales in parallel. This module was included so the network can further analyze the image in its different scale steps and make it able to take spatial information at different fields of view for each downsampling step. These modules are based on the original Scene parsing module [6] which was designed with the idea of analyzing the inputs at different scales in parallel to better incorporate more contextual information. The full network architecture of the PSPU-net can be seen in Fig. 3. The final output of the network is a feature map of 4 channels representing the same information as described for the 3D U-net. The total number of trainable parameters was of 30.83 million, and the neural network occupied 362 MB.

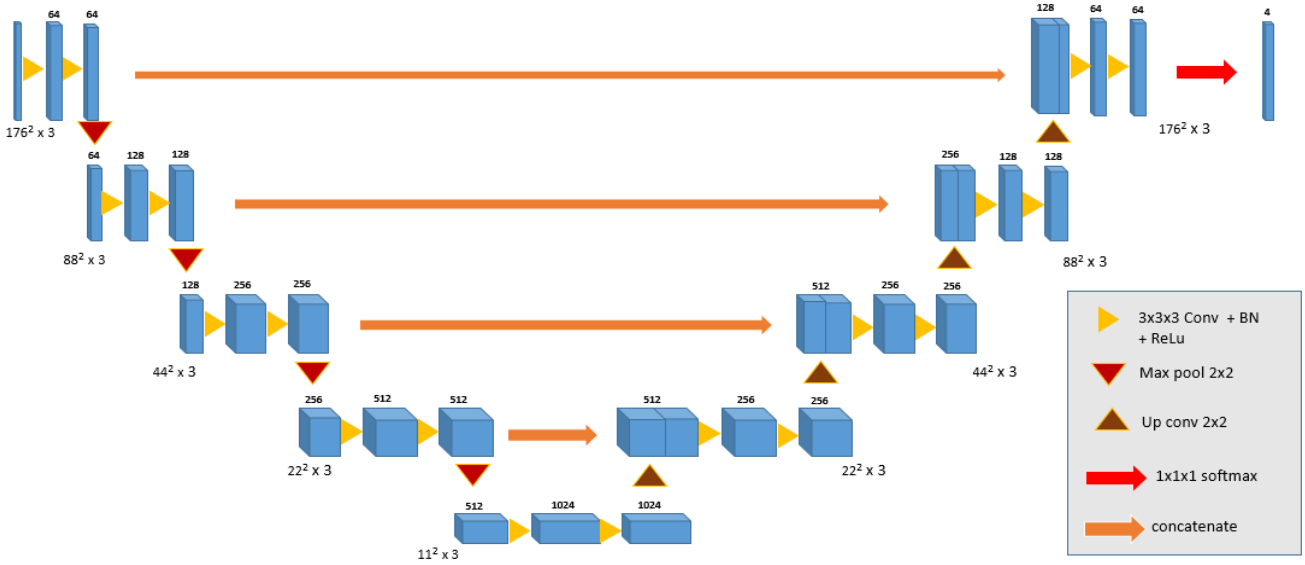

Fig. 1 3D U-net architecture design

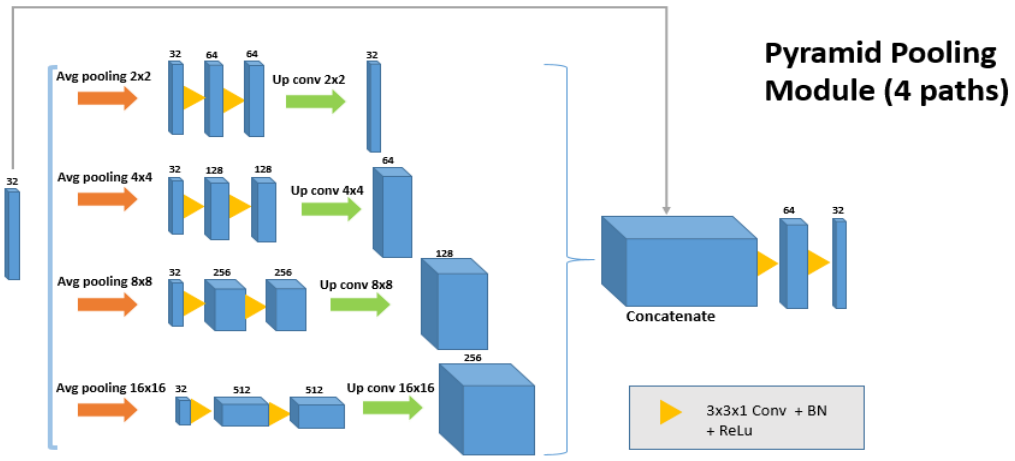

Fig. 2 Four paths PSP block design

D. *Network training*

Both networks were implemented by tensorflow 2.1 (www.tensorflow.org, Google Brain, Mountain View, California, U.S.) using its Keras API. The hardware used comprised a GPU RTX 2080 Ti with 11Gb of RAM (Nvidia Corporation, Santa Clara, California, U.S.) CPU i9 9900K (3.6 GHz) and 64 GB of RAM, running on Windows 10 operating system. Both networks were trained for 50 epochs using both the training dataset and validation dataset. After some testing the training was set up using ADAM optimizer with a learning rate of 0.001 using a batch size of 3. These hyperparameters obtained the best performance in both networks. These values match those described in similar neural networks for the same tasks [9, 10, 11].

The weighted generalized Dice loss [12] was used as loss function. The weights assigned to each label were predefined as 0.1 for background and 0.3 to the different tissues. With this we assigned a low weight to the background to account for possible unbalanced label presence and at the same time the remaining weights assured that the same importance was assigned to the three tissue labels. The training dataset included 260 cases resulting in a total of 2045 volumes of 3 slices (inputs) and the validation 40 cases resulting in 319 volumes of 3 slices. The training dataset was incremented through data augmentation in order to help prevent overfitting. To do this the inputs were randomly selected and modified applying a random rotation (angles between -30° and 30°), a random zoom factor (between 0 and 0.1) and a random shear (angles between -20° and 20°) to generate a total of 6295 extra inputs, resulting in a total training dataset of 8340 inputs for the training. All the transformations were only applied in the xy plane.

III. Results

To validate the results of the segmentations offered by both networks against the manual segmentations we calculated the dice scores (global dice and the dice for each individual tissue) and the absolute relative volume error derived from the segmentations. Our interest was focused on the diastolic frames of the cine acquisitions. For this we employed the remaining 99 cases of our database. Additionally, we also applied the networks on systolic frames to determine how well the networks performed against images with similar features but of different nature. In the latter case we only compared against the RV and LV cavities as these

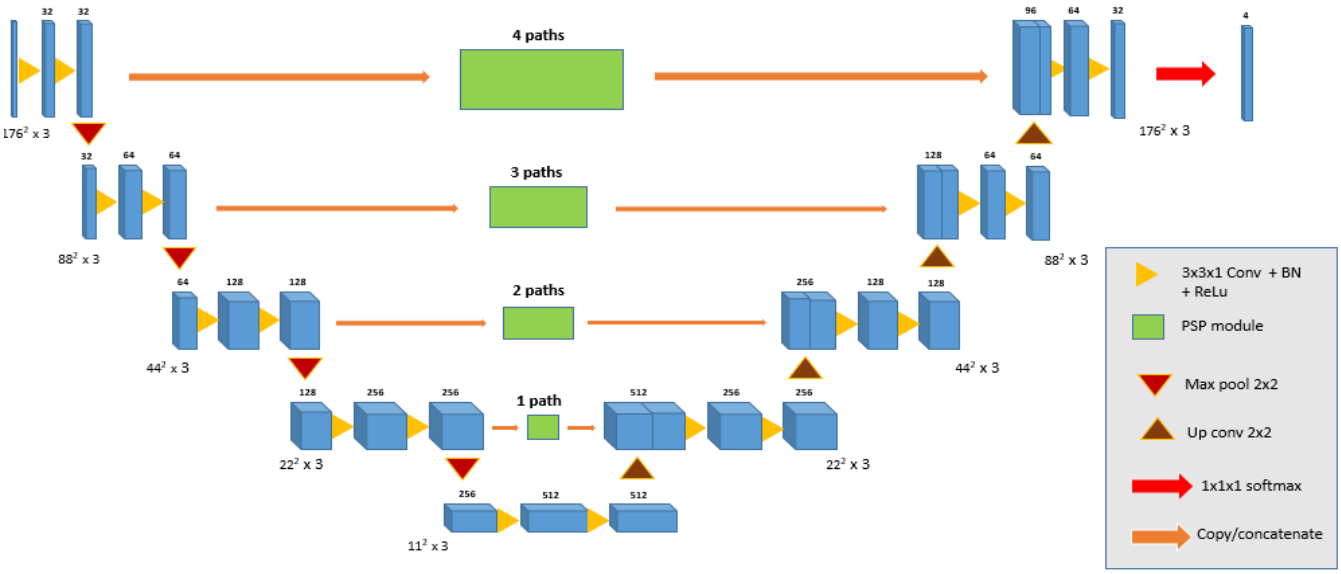

Fig. 3 PSPU-net architecture design

frames did not include the LV cavity segmentation. For this we employed a total of 98 cases.

All the results and statistics refer to the segmentation results of the whole image acquisition. The networks work only with volumes of 3 slices. To apply the segmentations, we sampled sub-volumes from the whole cine acquisition. Afterward we combined the segmentation results to obtain the whole volume segmentation. This process is schematized in Fig. 4.

A. *Network training performance*

The 3D U-net required a total of 27 hours to complete training while the PSPU-net needed 20 hours. The training history of the loss function for both networks is presented in Fig. 5, it can be seen as how both performed similarly, with similar training and validation loss evolutions. It is also noticeable how in both cases after epoch 4 the validation loss stays fluctuating around 0.09 while the training loss keeps decreasing. This means that both networks were capable of reaching their optimal status in a small amount of time and afterwards they entered in an overfitted status that did not decrease its accuracy (validation loss stays approximately the same while training loss keeps decreasing). Another important fact is that the training in the PSPU-net fluctuated less in terms of its loss function. In the case of the 3D U-net there seems to appear more little fluctuations. This could mean that the PSPU-net is slightly more stable in its training process.

B. *Segmentation results*

Tables 2 and 3 show the results for the segmentation quality of the tissues with both networks in diastolic and systolic frames respectively. The tables present the median and standard deviation values. It can be seen that both

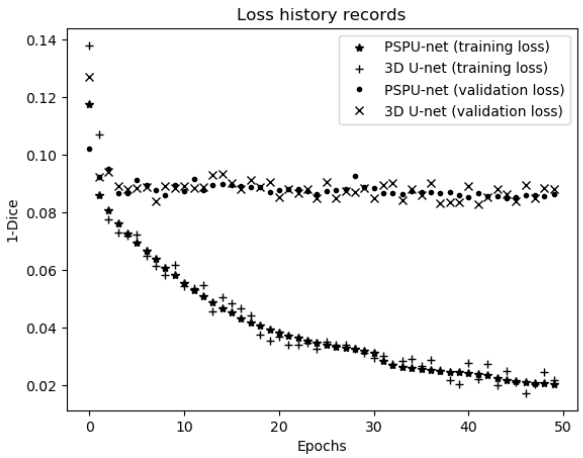

Fig. 5 Training history of the loss function (1-Dice) for both networks.

networks perform similarly in the diastolic frame (which was the frame on which they were trained on). Overall it seems that the PSPU-net offers slightly better dice scores and more reduced standard deviations, which could mean that the PSPU-net is more robust. In both networks the best results are those of the LV cavity (median score of 0.95), followed by the RV cavity (median scores of 0.9) and the LV myocardium having the worst scores (median scores of 0.875). Fig. 6 presents slices of a case with an overall dice score close to the average. With this example it can be seen that the segmentations offer satisfactory results. The results obtained for the systolic frames show a similar trend. In this case the superiority of the PSPU-net is more clear, with both superior dice scores and reduced standard deviations, making the PSPU-net a more robust network. Also it should be noted that for the LV cavity both networks perform well enough to be

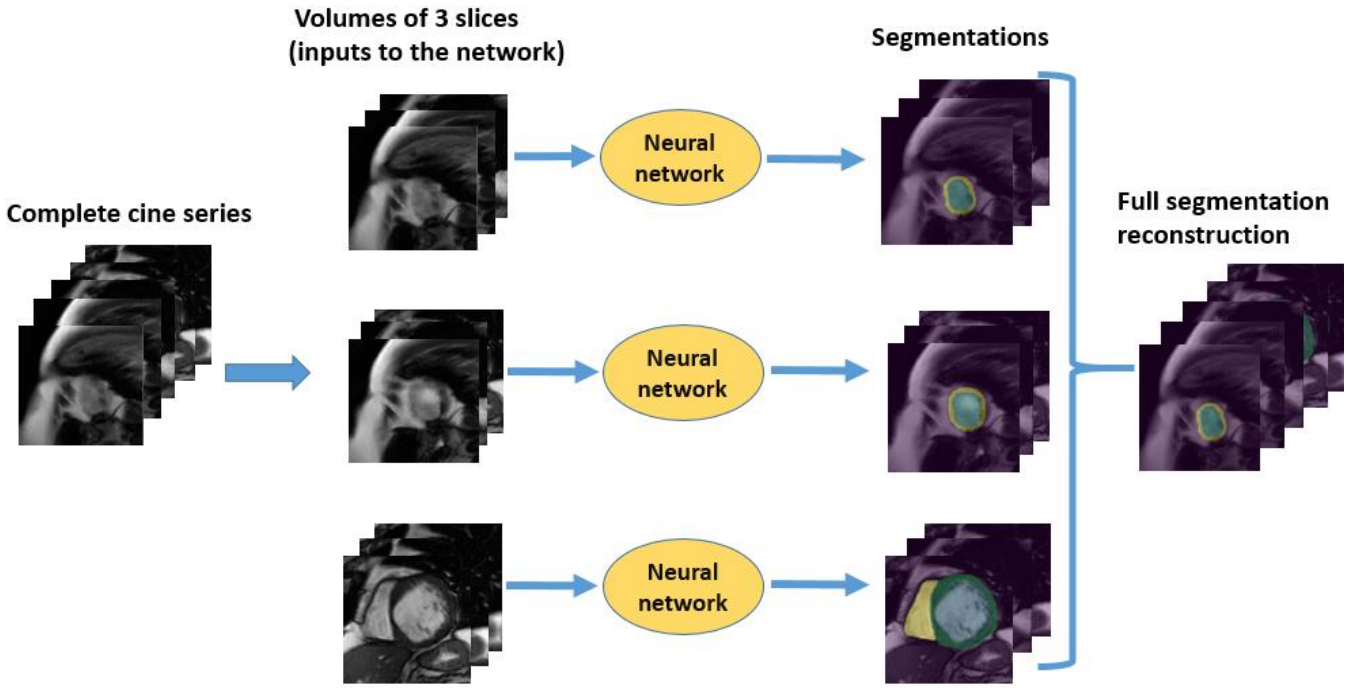

Fig. 4 Inference method for whole volume segmentation

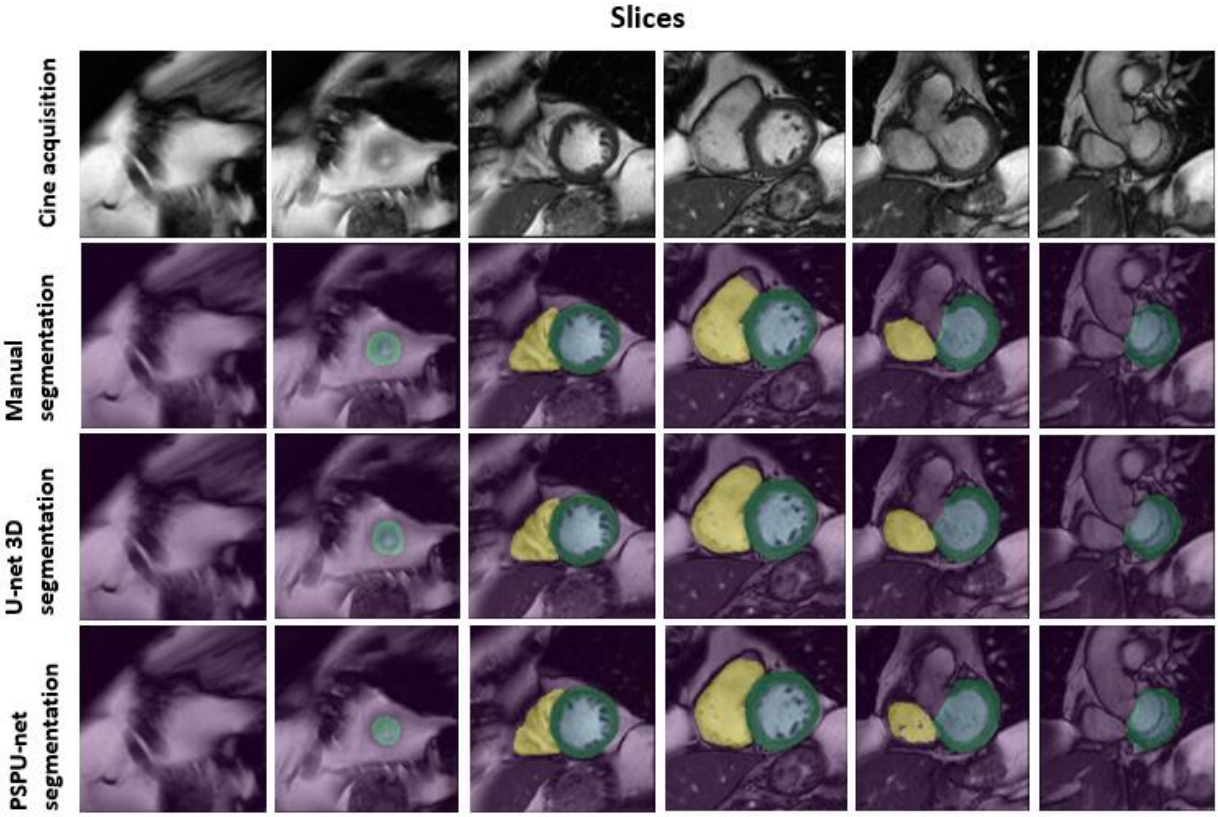

Fig. 2 Slice examples of a representative case of the segmentation offered by both networks and manual segmentation. Blue: Left ventricle cavity; green: left ventricle myocardium; yellow: right ventricle cavity.

considered of practical use (median dice scores of 0.88 and 0.9). In the case of the RV cavity the quality is notably reduced in both networks (median dice scores of 0.78 and 0.8).

With respect to the performance in the inference process using our hardware, we obtained a mean processing time of 0.91 seconds for the PSPU-net and 1.13 seconds with the 3D U-net when the cases are segmented individually. Although we did not test it, taking into account that the GPU parallelizing capabilities allow for the inference to be done in bigger batches (we employed batches of 1) we can assume that these times can be reduced when applying multiple segmentations at the same time.

C. Volume estimation results

The volume estimation metrics are presented in tables 4 and 5. The results show the absolute relative error in volume estimation for the different tissues. Similarly, to the dice scores, in this case there are not big differences between networks in the diastole. However, it is noticeable that the PSPU-net performs better in the LV myocardium and RV cavity. In the LV cavity the relative error is lower with the 3D U-net but in a much lower scale.

TABLE 2. Dice metric of the segmentations obtained in the diastole test set (99 cases). Median and standard deviation values.

Neural network	Whole segmentation	LV cavity	LV myocardium	RV cavity
3D U-net	0.907±0.028	0.956±0.021	0.875±0.039	0.904±0.042
PSPU-net	0.910±0.026	0.955±0.021	0.875±0.037	0.905±0.036

TABLE 3. Dice metric of the segmentations obtained in the full systole set (98 cases). Median and standard deviation values.

Neural network	Whole segmentation	LV cavity	RV cavity
3D U-net	0.826±0.067	0.881±0.073	0.781±0.085
PSPU-net	0.848±0.053	0.896±0.053	0.798±0.073

TABLE 4. Absolute relative error of the volume estimation obtained with the segmentations in the diastole test set (99 cases). Median and standard deviation values.

Neural network	LV cavity	LV myocardium	RV cavity
3D U-net	0.025±0.032	0.048±0.049	0.058±0.070
PSPU-net	0.026±0.033	0.039±0.051	0.051±0.047

TABLE 5. Absolute relative error of the volume estimation obtained with the segmentations in the full systole set (98 cases). Median and standard deviation values.

Neural network	LV cavity	RV cavity
3D U-net	0.115±0.121	0.258±0.221
PSPU-net	0.084±0.118	0.234±0.196

In the case of the systolic frames the results show a similar tendency to that seen with the dice scores, the PSPU-net performs noticeable better. Still the volume estimation error is

big in both networks, which makes them not suitable to be used in the systole time frame for volume estimation.

IV. Conclusions

We have presented a novel version of the U-net that employs PSP blocks for the segmentation of short-axis diastole frames of MRI cine-sequences and compared its results with a standard version of a 3D U-net. The results for both segmentation and final volume estimation of the tissues showed that the PSPU-net offered high quality results, slightly superior to those of the 3D U-net with greater stability in the results.

We additionally tested both networks with systolic frames to see how well they could perform with images with similar features but different nature. Comparing the performance of both networks in a different setting helped to determine which network was generalizing the calculation of image features better. The fact that the PSPU-net has notably better results in the systolic frames could mean that the features it is capable to extract work globally better and are more robust to outliers to those extracted by the 3D U-net. Adding to this the fact the PSPU-net is a smaller network (about 3 times smaller) makes it also a more efficient one.

We conclude that the PSPU-net is a convolutional neural network for segmentation that can offer high quality results in the task analyzed and that adding PSP blocks in U-net architectures can help in improving the robustness of these architectures.

Acknowledgment

DM acknowledges financial support from the Conselleria d'Educació, Investigació, Cultura i Esport, Generalitat Valenciana (grant AEST/2019/037), from the Agencia Valenciana de la Innovación, Generalitat Valenciana (ref. INNCAD00/19/085), and from the Centro para el Desarrollo Tecnológico Industrial (Programa Eurostars-2, actuación Interempresas Internacional), Ministerio de Ciencia, Innovación y Universidades (ref. CIIP-20192020). We are grateful to Andrés Larroza for his valuable technical assistance in the project.

References

[1] M. Nichols, N. Townsend, P. Scarborough and M. Rayne, "Cardiovascular disease in Europe 2014: epidemiological update", European heart journal, vol 35, no. 42, pp. 2950-2959, 2014.

[2] C. F. Baumgartner, L. M. Koch, M. Ollefeys and E. Konukoglu, "An exploration of 2D and 3D deep learning techniques for cardiac MR image segmentation", in International Workshop on Statistical Atlases and Computational Models of the Heart, Springer, Cham, Sep. 2017, pp. 111-119.

[3] O. Bernard et al, "Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved?", IEEE transactions on medical imaging, vol 37, no. 11, pp. 2514-2525, 2018.

[4] Q. Tao, B. P. Lelieveldt and R. J. van der Geest, "Deep Learning for Quantitative Cardiac MRI", American Journal of Roentgenology, vol 214, no. 3, pp. 529-535, 2020.

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

[6] H. Zhao, J. Shi, X. Qi, X. Wang and J. Jia, "Pyramid scene parsing network", in Proceedings of the IEEE conference on computer vision and pattern recognition, Honolulu, HI, USA, Jul. 2017, pp. 2881-2890.

[7] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox and O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation", In International conference on medical image computing and computer-assisted intervention, Springer, Cham, Oct. 2016, pp. 424-432.

[8] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift", arXiv preprint arXiv:1502.03167, 2015.

[9] A Q. Tong et al, "RIANet: Recurrent interleaved attention network for cardiac MRI segmentation", Computers in biology and medicine, vol. 109, pp 290-302, 2019.

[10] E. Abdelmaguid et al, "Left ventricle segmentation and volume estimation on cardiac mri using deep learning", arXiv preprint arXiv:1809.06247, 2018.

[11] A. H. Curiale, F. D. Colavecchia and G. Mato," Automatic quantification of the LV function and mass: a deep learning approach for cardiovascular MRI", Computer methods and programs in biomedicine, vol. 169, pp. 37-50, 2019.

[12] A C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin ad M. J. Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations", In Deep learning in medical image analysis and multimodal learning for clinical decision support, Springer, Cham, 2017, pp. 240-248.