Computer Methods and Programs in Biomedicine 208 (2021) 106275

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Automatic left ventricle volume calculation with explainability through a deep learning weak-supervision methodology

Manuel Pérez-Pelegrí[a], José V. Monmeneu[b], María P. López-Lereu[b], Lucía Pérez-Pelegrí[c], Alicia M. Maceira[b], Vicente Bodí[d,e], David Moratal[a,*]

[a] Center for Biomaterials and Tissue Engineering, Universitat Politècnica de València, Camí de Vera, s/n, 46022 Valencia, Spain
[b] Unidad de Imagen Cardíaca, ERESA-ASCIRES Grupo Biomédico, Valencia, Spain
[c] Facultad de Enfermería, Universidad Católica de Valencia San Vicente Mártir, Valencia, Spain
[d] Departamento de Medicina, Universitat de València, Estudi General, Valencia, Spain
[e] Servicio de Cardiología, Hospital Clínico Universitario de Valencia, INCLIVA, CIBERCV, Valencia, Spain

ARTICLE INFO

Article history:
Received 10 March 2021
Accepted 2 July 2021

Keywords:
Magnetic resonance imaging
Deep learning
Left ventricle
Weak supervision
Explainability
Segmentation

ABSTRACT

Background and objective: Magnetic resonance imaging is the most reliable imaging technique to assess the heart. More specifically there is great importance in the analysis of the left ventricle, as the main pathologies directly affect this region. In order to characterize the left ventricle, it is necessary to extract its volume. In this work we present a neural network architecture that is capable of directly estimating the left ventricle volume in short axis cine Magnetic Resonance Imaging in the end-diastolic frame and provide a segmentation of the region which is the basis of the volume calculation, thus offering explainability to the estimated value.

Methods: The network was designed to directly target the volumes to estimate, not requiring any labeled segmentation on the images. The network was based on a 3D U-net with extra layers defined in a scanning module that learned features like the circularity of the objects and the volumes to estimate in a weakly-supervised manner. The only targets defined were the left ventricle volumes and the circularity of the object detected through the estimation of the π value derived from its shape. We had access to 397 cases corresponding to 397 different subjects. We randomly selected 98 cases to use as test set.

Results: The results show a good match between the real and estimated volumes in the test set, with a mean relative error of 8% and a mean absolute error of 9.12 ml with a Pearson correlation coefficient of 0.95. The derived segmentations obtained by the network achieved Dice coefficients with a mean value of 0.79.

Conclusions: The proposed method is capable of obtaining the left ventricle volume biomarker in the end-diastole and offer an explanation of how it obtains the result in the form of a segmentation mask without the need of segmentation labels to train the algorithm, making it a potentially more trustworthy method for clinicians and a way to train neural networks more easily when segmentation labels are not readily available.

1. Introduction

Correct assessment of the left ventricle (LV) is critical in the diagnosis of cardiovascular diseases, which are one of the current major health problems in advanced countries [1]. Magnetic resonance imaging (MRI) is one the most used diagnostic tools for cardiac structure and function assessment. The characterization of the LV is usually done through the volume calculation in the end-systolic and the end-diastolic time frames, from which the volume, mass and ejection fraction of the LV can be derived. However, obtaining these parameters is often a time consuming task, a clinical user will require the segmentation of the LV in the end-systolic and end-diastolic frames to be done either manually or semi-automatically with the help of specialized software in order to extract the volumes.

* Corresponding author.
E-mail address: dmoratal@eln.upv.es (D. Moratal).

https://doi.org/10.1016/j.cmpb.2021.106275

In this context, deep learning techniques have been used to assess this problem. Most of the work has focused in using segmentation procedures before obtaining the volumes for specific datasets obtained from a single machine source [2–5] and a thorough ongoing research still goes in this direction aiming to accurately segment the left ventricle in multi-center settings [6,7]. However, the exploration of biomarkers extraction from medical images [8] has also been studied using regression networks including age estimation from T1 MRI brain scans [9], diabetic retinopathy detection in retinal fundus images [10], morphometric parameters of the corneal endothelium (cell density, cell size variation, and hexagonality) in corneal endothelium microscopy images [11], skin disease detection and classification from skin lesion photographs [12], detection of osteoporosis and emphysema on chest CT [13] or Agatston score obtained from chest CT scans of the heart [14]. In these works, a prior manual segmentation of the region from which the biomarker is obtained was not needed. In these cases, the neural network is trained to automatically predict the biomarker value from the image. These types of networks work using a convolutional neural network that learn to extract relevant features followed by fully connected layers that use these extracted features to give the final biomarker prediction. These algorithms have the disadvantages that they require large amounts of data (in the order of thousands) [9,13,14] and are often viewed as "black boxes" where explainability of how the network obtain the result is hard. Explainability is still an ongoing part of research in deep learning and is one of the most important factors to solve if a major trust is to be gained for this field in the clinical setting [15,16].

In this work we propose a neural network architecture for direct volume estimation of the LV in the end-diastole frame, but the design offers the capability of offering explainability of the results by indirectly being able to derive a segmentation of the region of the LV where the network has based its final output. Achieving this required the investigation of weak-supervision techniques [17] which are employed for segmentation when the segmented ground-truth is not available for training, and instead other type of indirect information is used in the training procedure. This approach has been recently taken with good results in other medical imaging contexts as stated in [18] were they could obtain pectoralis muscle area (PMA), subcutaneous fat area (SFA) and liver mass area in single slice computed tomography (CT), and Agatston score estimated from non-contrast thoracic CT images (CAC) without training for the specific target. This weak-supervision methodology has been employed using different features to train with, like labels of the image [19,20], seed points of the region to segment [19], regions of interest [21–23], or points around the contour of the region to segment [24]. In our case we are simply using a specific biomarker (the volume of the region) instead of elements that already helps to determine the location of the objects, which is a novel approach in the weak-supervision field. Additionally, we encode information regarding the fact that the LV is a circular object in the 2D plane to help the network learn to identify objects with this property.

2. Previous research

Most of the work done in regards to applying convolutional neural networks to the problem at hand aim to obtain the segmentation of the LV using previously segmented cases to train them. This approach has demonstrated to offer state of the art results for specific datasets [2–5]. The same methodology has been explored with multi-center/multi-vendor source image datasets, in [7] the authors tested a neural network for a test set coming from multiple scanners after being trained with a dataset from a single scanner and in [6] the authors designed a neural network trained on different scanner images. All these works used either U-net architectures or variations of it.

Recently, other methods that do not employ neural networks and that are capable of obtaining state of the art results in segmentation have also been described. Particularly in [25,26] an algorithm employing firstly a ROI selection of the LV region followed by a slope difference distribution threshold (SDD) and circular Hough transform has been applied to the problem of segmentation and LV detection successfully. This method is available in MATLAB code at https://uk.mathworks.com/matlabcentral/fileexchange/78417-sdd-lv-segmentation-for-comparison-with-dl-and-cnn-methods?s_tid=prof_contriblnk.

In the case of the LV direct volume regression with neural networks, there is a lack of work compared to the problem of segmentation and few studies have addressed this problem. In [27] the authors applied a convolutional regression network for the LV volume estimation in both end-systolic and end-diastolic frames with a large dataset with 1140 subjects (Data Science Bowl Cardiac Challenge Data) using 5 convolutional layers followed with 3 fully connected layers. A similar approach was taken in [28] using the same dataset comprising 1140 subjects. In this case, the authors preprocessed the images in order to crop only the region of interest that included the LV in a ROI and then fed this data to a neural network comprising 13 convolutional layers followed by 3 fully connected layers

3. Materials and methods

3.1. Image dataset

Our dataset consisted of 397 short-axis stacks of MRI covering both the left and right ventricles obtained from the same MR scanner. This dataset comprised a total of 397 different patients (270 men and 127 women), with an age of 64.53 $\pm$ 12.35 years old (63.27 $\pm$ 11.98 years old for men, and 67.42 $\pm$ 12.75 years old for women) (mean $\pm$ standard deviation). The diagnosis for the patients included a great variety of conditions like presence of fibrosis, necrosis, ischemia, functional affection of LV (ejection fraction lower than normal and/or affected segmental contractility), or "healthy" subjects (subjects with no cardiac pathology associated). The final diagnosis and the number of cases for each condition is presented in Table 1. For the experiments we used only the time frame corresponding to the 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 varied across the cases, ranging from 0.57 $\times$ 0.57 mm^2 to 1.09 $\times$ 1.09 mm^2. The slice thickness and spacing between slices was constant in all cases, 7 mm and 3 mm respectively. The resulting image sizes varied from 144 $\times$ 144 to 256 $\times$ 256 and the number of slices ranged from 8 to 14. All the images were resampled using bilinear interpolation to a constant in-plane spatial resolution of 2 mm^2 with an image size of 88 $\times$ 88. This downsampling was applied in order to reduce the number of features (represented by each voxel) for the network to process. Training was tested with original sizes but after several experiments it was seen that reducing the sizes improved the training performance. The z-axis was left untouched in the resampling process. Additionally, the 3D stacks were normalized to make the pixel values range from 0 to 1 using min-max normalization.

Every case of the entire dataset was categorized in one of the 11 categories corresponding to the diagnosis category (see Table 1). This was done in order to ensure that the split between training, validation and test sets had similar distribution with respect to the

Table 1
Classification of the dataset in categories according to its clinical diagnosis

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	135
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

diagnosis. Finally, the dataset was randomly split in training (259 cases, 65%), validation (40 cases, 10%) and test set (98 cases, 25%). Additionally, we had access to the manual segmentations of the LV performed by mutual consensus of two cardiologists with more than 10 years of experience. The volumes used as inputs to the network had been previously derived from these manual segmentations.

3.2. Neural network architecture

The neural network designed is based on the classic U-net [29], more specifically it is a 3D version of it [30] with some notable changes that we proceed to describe. The network takes as inputs images of 88 × 88 × n. The value of n can be variable and it represents the number of slices within the input volume (which in our dataset ranged from 8 to 14). All the layers include batch normalization (BN) [31] to improve the training performance. Every convolution in the network is a 3 × 3 × 3 convolution and has a rectified linear unit as activation function (ReLU). To reduce the feature maps (also called channels) spatial size maxpooling operations of size 2 × 2 × 1 are applied to halve the xy plane, while preserving the number of slices of the feature maps. The "bottleneck" layer output is regularized with L1 activity regularization to force it to encode meaningful features and set to 0 non-important ones. The up-sampling path is composed of up-convolutions of size 2 × 2 × 1 to recover the xy size, the up-convolutions are a way to upsample inputs to specific sizes with the advantage that the upsampling process is associated to a convolution with weights that can be learned, and thus optimize the upsampling of the feature maps. Concatenation of the final feature maps obtained in the contracting path are also applied to the feature maps in the upsampling path after every up-convolution in the same manner as the U-net, this helps the network to recover the original spatial resolution that was reduced along the contracting path [29]. The final layer is a 1 × 1 × 1 sigmoid activation function that gives the probability of each voxel being part of the left ventricle region. This layer has also L1 activity regularization to force that only the LV appears in the class activation map (CAM). This layer outputs the CAM, that gives a probability map for the voxel classification [32,33], from which the network will derive the final output volume and indirectly a segmentation of the region of interest. The full network architecture can be seen in Fig. 1 and an example of a CAM output is presented in Fig. 2. The final layers of the network correspond to the scanning module that apply sweeps to the CAM in order to derive meaningful parameters, mainly the final volume and the diameter of the detected object. In order to obtain the volume, we incorporated a non-trainable 25 × 25 × 1 convolution filled with ones, after which a max pooling layer (88 × 88 × 1) extracts the maximum for each slice and then applies a sum along the third dimension. The non-trainable convolution size was defined based on the usual LV size in the slices where it appears bigger in the downsampled images used as inputs. This specific size ensure that the LV can fit with a small margin in it. The maxpooling ensures that only the biggest object is taken into account for the volume calculation, as in rare cases the CAM can produce other small residual objects of high probability. The second scanning incorporates 2 non-trainable convolutions of ones of size 1 × 88 × 1 and 88 × 1 × 1 and the output is then averaged. With this scanning the net is forced to detect the largest diameter of the objects present in the net along the slices. We then combine these with the first scanning convolution (which outputs the areas for every slice) in order to output the relation between the two, which should be close to the number π (as the left ventricle is approximately circular). This circularity feature extractor is what ultimately allows the net to detect circular objects whose volume match that of the target, which corresponds to the LV. Fig. 3 show a schematic of the scanning layers described.

3.3. Network training

The network was implement using tensorflow 2.1 (www.tensorflow.org, Google Brain, Google, LLC, Mountain View, CA) using its Keras API. The hardware employed included a PC computer with an Intel Core i9 9900k 3.6 GHz CPU (Intel Corporation, Santa Clara, CA), 64 GB of RAM, a GPU RTX 2080 Ti with 11 GB of RAM (Nvidia Corporation, Santa Clara, CA) running on Windows 10 operating system (Microsoft, Redmond, WA). The network was trained for a total of 50 epochs using both the training and validation dataset. After running some test, the training was set up with ADAM optimizer with a learning rate of 0.001 with a batch size of 5.

During the training we used a custom loss function that took into account the L1 regularization terms, and the mean absolute errors of the estimated volumes and the derived π value. Due to the nature of each error and the regularization terms, there was a significant difference in the scale of the contributions of each term to the loss. Specifically, L1 regularization penalizes the signal in the feature maps of the bottleneck and the CAM (which can be very high reaching the order of tens of thousands), the π error can be very low in contrast (in the order of units), and the LV volumes can reach values in the order of hundreds (using milliliters measurements). To compensate for this, we multiplied the error of π by 100 and the L1 regularizations were weighted by 10^{-3}. These specific weights were chosen based on the observed errors during the network training after several experiments. The values that optimized the training process were the ones described, which means that an equilibrated contribution was required for a good performance. The final loss function was the sum of all these parameters. This loss function can be checked in Eq. (1), where MAE (volume) and MAE (π) represent the mean absolute error for the volume prediction in milliliters and for the value of π estimated

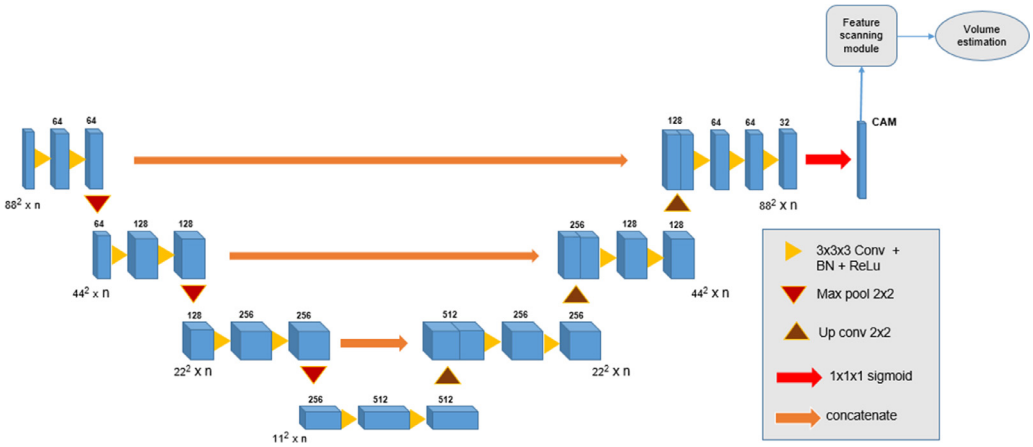

Fig. 1. Neural network architecture design. The design is similar to that of the 3D U-net, but the last layer of the bottleneck and the CAM layer include L1 activity regularization. The CAM layer is then passed to the features scanning module which produces the final volume estimation. Above every block is the number of feature maps associated to it, and under it at the side of each stage is the size of the feature maps, where the squared number refers to *xy* plane size and *n* the *z* size. The number of feature maps generated start at 64 and gets doubled at each pooling stage in the contracting path. Conversely, along the upsampling path the number of feature channels is halved at each step, until reaching a final convolution that outputs a 32-channel block before the last one that generates the CAM.

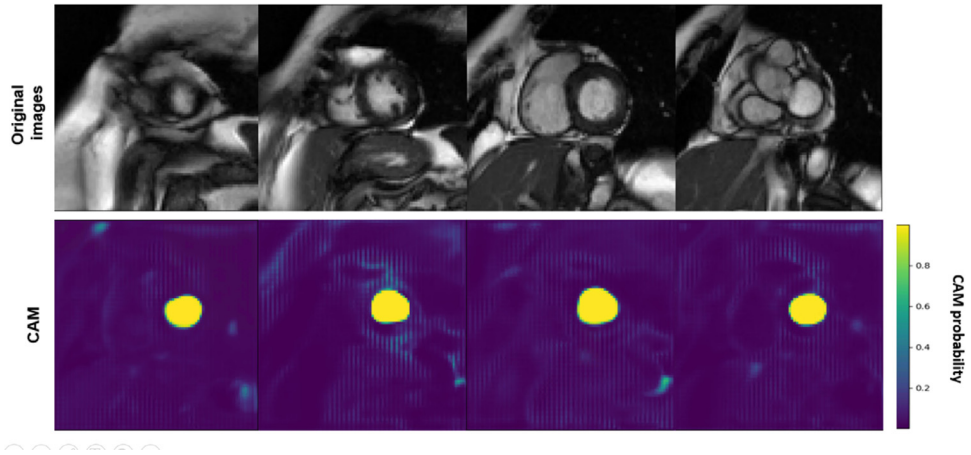

Fig. 2. Example of the output obtained by the CAM layer of the proposed neural network. It can be seen how the left ventricle region corresponds to a high probability in the CAM. *CAM: class activation ap.*

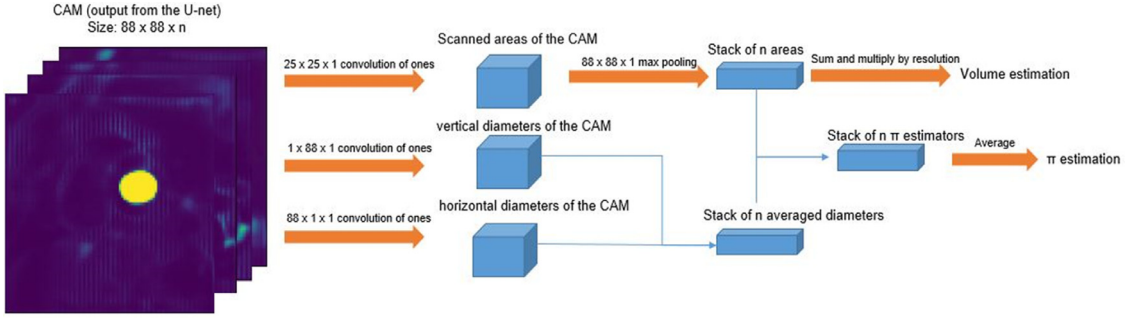

Fig. 3. Scanning module designed to extract two features from the CAM. The first convolution scan obtains the volume of the biggest region. Then a second scanning of convolutions finds the average diameters for each slice, which combined with the previous scanned areas output an estimation of the number π in order to help the network focus on circular objects.

respectively.

$$Loss = MAE(volume) + 100 \times MAE(\pi) + 10^{-3}(L1(bottleneck) + L1(CAM)) \quad (1)$$

The training dataset was increased through data augmentation techniques. Specifically, for each batch we added an extra batch during training. This additional batch was obtained from random samples from the training dataset that applied random rotation (between -30° and +30°), random shear (between -20° and +20°), random translation (between -15 and +15 pixels) and added gaussian noise (mean of 0.035 and standard deviation of 0.01).

4. Results

In this section we present the results offered by the network. We analyzed two types of results in the test set. One was the direct

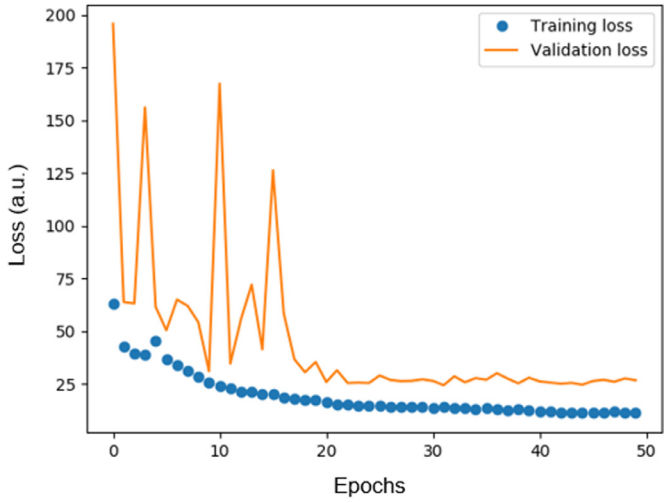

Fig. 4. Training and validation history of the network. The training loss kept decreasing in a stable manner, with little improvement from epoch 30. The validation loss took some time to stabilize until it reached epoch 20 with very high losses and fluctuations before reaching this point. This shows that in the initial stages the training was not obtaining meaningful features and that the training of such network can take some time and is of high difficulty. *a.u.: arbitrary units.*

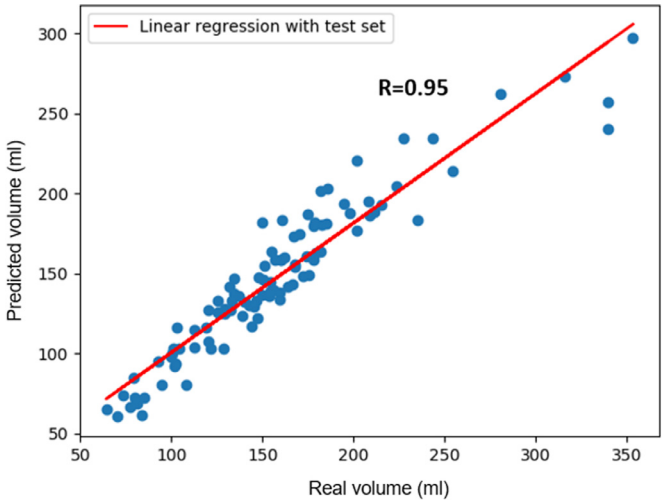

Fig. 5. Cloud of points representing the real volumes against the predicted volumes from the network for the test set. The analysis of the results shows a high correlation (R=0.95) and the values closely match in most cases. The regression showed a slope of 0.81 and bias of 19.41. Although the correlation is big, it is noticeable that with the cases of higher volume values there is an underestimation that seems to be more apparent from 250 ml.

volume estimation offered by the network and the other was the segmentation derived from the CAM employed by the network to obtain the volume. Furthermore, we analyze the performance of the training process.

4.1. Training performance

The neural network was trained for 50 epochs and needed 12 hours to complete the training. The training history of the loss function can be seen in Fig. 4. The training loss improved during all the 50 epochs, however the validation loss shows a more erratic evolution in the first steps until it stabilizes in epoch 20 and stays approximately constant during the rest of the training. This could mean that the training of this type of network is complicated, as it takes time to actually start learning meaningful features to get good results in the validation dataset without overfitting the training set.

4.2. Volume estimation results

The distribution of the relative errors of the volume estimation from the network followed a mean and standard deviation of 8.50 $\pm$ 6.60 % with the majority of the cases falling between 0 and 10 % relative error. In terms of the absolute error the distribution followed a mean and standard deviation of 9.127 $\pm$ 18.888 milliliters. After exploring the results, we encountered a slight tendency to underestimate the volumes. Fig. 5 present the cloud of points of the real volumes against the estimated volumes in milliliters. It shows a high correlation (R=0.95) and it also shows how the tendency to underestimate the volume is greater when the LV is bigger. This trend starts to be more apparent at 250 ml, however there were few cases that surpassed this limit (6 cases in total, 6.12% of the test set).

As for the computation time required by the network to output its prediction, using the hardware described the 98 cases of the test set requires a total of 104.85 s to complete using a batch of 1 (we used a batch of 1 in order to check the time required for independent computations), resulting in an average of 1.07 s per case.

4.3. Derived segmentation results

An analysis of the masks derived from the CAM layer was also performed to ensure that the estimated volumes were in fact obtained from the corresponding region of the LV, which was the aim in order to offer the desired explainability. For this analysis we employed the Dice coefficient [34] which measures the degree of overlap between the estimated object and the manual segmentation. In order to generate the masks, we performed two steps: first we applied a threshold of 0.9 to the CAM and then we retrieved only the biggest object present (which is the one that the network used to derive the volume) as in some rare cases small residual objects outside the LV appeared in the CAM. Fig. 6 shows a representative example of the segmentation achieved with this procedure. After exploring the segmentations, we encountered two tendencies: they tended to overestimate the mask at the apical region and to slightly underestimate the remaining slices. This matches the results of the volumes estimated directly by the network, leading to a slight underestimation of the final volume. The location of the LV was correct in all cases.

Based on these findings we additionally applied a simple post-processing to the masks in order to improve the segmentation: elimination of the more apical slice followed by a dilation of $5 \times 5 \times 1$ pixels at each slice. These improved the masks significantly. The Dice coefficient distributions for both the basic masks derived from the CAM and the masks with the post-processing followed a mean and standard deviation of 0.720 $\pm$ 0.053 and 0.791 $\pm$ 0.042 respectively. These results show that applying the post-processing described resulted in better segmentations, which confirms the tendency observed in the masks directly derived by the CAM.

5. Discussion

We have presented a fully convolutional neural network that is capable of obtaining the LV volume in the diastole time frame and derive a segmentation of the region to offer explainability. The network offers such outputs after being trained with only the actual volume values, not using any information regarding the location of the LV, but encoding information regarding the circular property of the LV in the images.

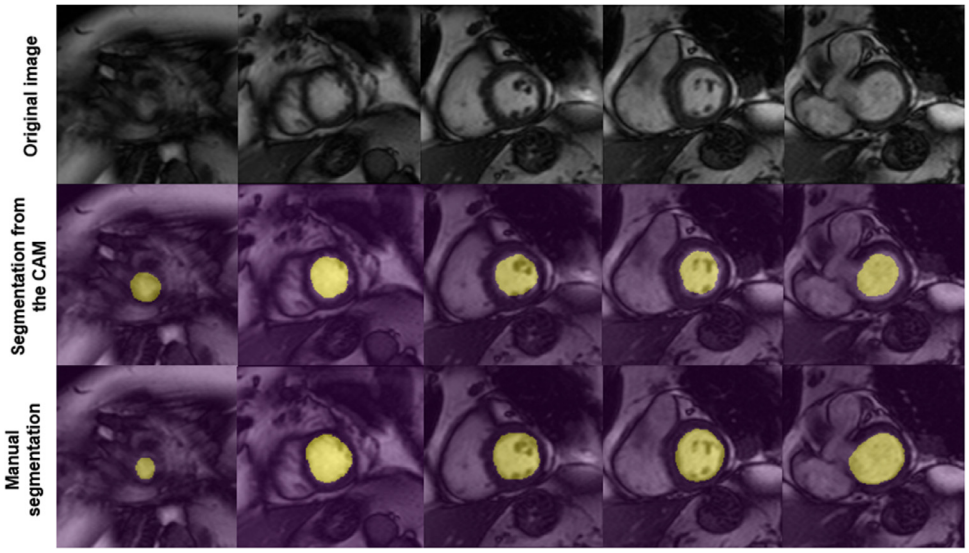

Fig. 6. Example of segmentations of the LV derived from the CAM in the test set. It is noticeable how at the apical slice there is a slight overestimation of the region, while in the remaining slices a small part of the LV is left out of the mask. Still, in all cases the mask location matched that of the LV.

The current tendency is to explore the use of convolutional neural networks to obtain the segmentation of the LV using previously segmented cases to train them. This type of work usually achieves state of the art result for specific datasets with reported Dice coefficients between 0.93 and 0.96 [2–5]. In the case of multicenter/multi-vendor studies, in [7] it was reported a Dice coefficient of 0.90 and in [6] the designed neural network achieved Dice coefficient values between 0.88 and 0.95 for the endocardium area depending on the region (apical, middle and base). All these results are promising and show very accurate matches. However, working with the pure segmentation methodology still poses the problem of the need to have such segmentations available. We have to take into account that in the clinical setting much of the cases are saved in databases were only the final output is saved [18] (LV mass, volume, ejection fraction, etc.), this is furthered supported as one of the biggest free-available dataset do not provide segmentation to work with (Data Science Bowl Cardiac Challenge Data, https://www.kaggle.com/c/second-annual-data-science-bowl/data), making access to segmentations a hard and time-consuming problem. Training the networks with only the volume values may give access to a wider spectrum of clinical databases were the value of the biomarker is usually readily accessible.

In regards to the method recently described in [25] the reported Dice scores 0.95 in the test set of the automated cardiac diagnosis challenge (ACDC) consisting of 50 subjects. These results show a similar quality as those of convolutional neural networks.

In contrast to the direct segmentation, the use of regression networks for direct biomarker estimation is a subject that has been employed extensively in the medical field for different topics but few work has been done for the LV volume biomarkers and its derivates.

Specifically, for the case of the LV direct volume regression with neural networks, few studies have addressed this problem. In [27] they reported good correlation results with values of 0.95 and 0.92 for the end-diastole and end-systole respectively using 337 cases as a test set. Additionally, they reported a mean error of 5.1 ml and 3.6 ml respectively. In [28] the reported results were worse with a mean error of 15.83 ml and 9.82 ml for the end-diastolic and end-systolic frames respectively using as test set a total of 440 cases. In these studies, they aimed to directly obtain the volume values of interest and they obtained reasonably good results in the predictions but this kind of approach lacked the ability of offering explainability to the result and how the network was obtaining the final volume and also required a large amount of cases to train the networks. These are problems that are not present in the case of segmentation networks, as they usually require a lower number of cases to train the networks and their output can be checked to ensure the correct assessment of the regions of interest and be trusted by clinicians.

The design of our neural network is different to the usual regression networks, in which usually a convolutional neural network learns to extracts the features from the image and then a fully connected neural network uses this features to obtain the biomarker estimation. This makes explainability even more hard to obtain, as the representation and dimensionality of the image is lost in the process. In contrast our neural network uses the typical U-net architecture whose output is designed to match the desired segmentation, and from the output obtained by the CAM layer it is capable of obtaining the volume of interest. We also had to encode through a scanning module the property of circularity of the object to segment in order to help the network determine the best region match. The final output to train the neural network is not the segmented region but the direct biomarker (the volume to estimate) but it also aims to obtain the segmentation of the region based on the CAM layer. This makes the design fall within the weak-supervision methodology, which is still a relatively novel and an ongoing research field within machine learning [15,16].

As exposed, one important consideration to take this kind of approach is that not every medical image database saves the segmentation obtained in the clinical setting, making recovery of such results difficult and time-consuming. In this setting weakly-supervised methods like the one proposed will be very useful in the near future as well, as the direct biomarkers are usually saved in the databases and are more widespread and easily available. Weakly-supervised methods have been normally used employing as target for the training some kind of information that already helps the neural network to encode spatial information or presence of the object of interest in some regard. Examples are found in [19,20] were they employed simply labels of the image to segment the object of interest, [19] used seed points within the region to train with and derive the full object segmentation, [21–23] used ROIs of the objects to segment as the target during training and [24] points spread around the contour of the region to segment. In contrast to all this works we used exclusively the actual biomarker

value (the volume) and incorporated some spatial information in the form of a circularity scanning layer using the prior knowledge of the natural shape of the LV.

Although the results offered by the network show a good correspondence with the real values, there are still some limitations to its capabilities, as the volume errors and the Dice coefficient measurements for the derived segmentation do not reach state of the art results offered by convolutional neural networks that are trained to directly segment the LV. We could achieve mean Dice coefficient values of 0.791 and a mean error value of 9.127 ml with a very high Pearson correlation of 0.95 for the estimated volumes for the 98 cases of the test set. These results are comparable to the ones obtained by regression networks and show a good match with the real values. Although these are promising results, they fall behind the state of the art results, specially from the neural networks that are trained to directly offer segmentations. We believe this is due to the limited number of cases used during training. We employed a total of 259 volumes with data augmentation techniques during training, but this type of problem offers better results with larger datasets in the order of thousands of cases. This could also be seen in the training process where it took a great part of the training time for the network to actually start learning meaningful features, as seen how the validation loss did not stabilize to an acceptable trend until reaching epoch 20. Still, the results in both the final output volume and the segmentation overlap show a good match indicating that the methodology employed for this problem is correct and that it could be improved and extended for the volume estimation in the LV in the end-systolic time frame and to additionally derive the ejection fraction. It is also important to note that we did not test the method on systolic frames due to the increased difficulty usually associated to it, as in this frame the LV is considerably smaller and we had access to a fewer number of samples for the systole to train the network model. Although we did not test it for the systolic frame some changes would probably be required, mainly the size reduction of the non-trainable convolution in the scanning layer to better fit the size of the LV in those frames. We believe the methodology could also be expanded to encompass other problems with similar aims in the medical imaging field, as the one of the main focus is in the segmentations obtained that aims to offer explainability to the final user.

Explainability in deep learning is a major concern in recent years due to the fact that this type of algorithms is difficult to interpret and are usually viewed as black boxes. This is especially important for neural networks that aim to extract direct biomarkers without using a segmentation of the region of interest (usually through regression networks), while traditionally biomarkers have been extracted after segmenting the region of interest. We believe that the use of the approach taken in this work will increase in the following years to come, as it offers the opportunity to broaden the use of more databases which lack segmentation information and aims to offer the biomarker result along a segmentation, which would provide the desired explanation of that result. This method works in contrast to current regression networks that provide the biomarker output without giving an explanation of how it was obtained by the network, making the output difficult to trust [15,16]. Specifically, in the clinical field this is crucial, as the decision-making is of utmost delicacy and will be considered in the legal setting in the near future for this type of algorithms [35].

6. Conclusions

We have presented a neural network design that achieves good results in the LV volume estimation, with values close to the real ones. Furthermore, the network allows the obtention of a mask of the LV derived from the CAM layer, which explains how the neural network estimates the volumes. The masks obtained also showed a good correspondence with the actual LV region, which ensured that the estimated volumes matched the LV. We think this methodology is important from a clinical perspective, since it allows for clinicians to understand how these algorithms work and explain the results instead of just offering the biomarker output leading to the widespread view of them working as black boxes. At the same time, it broadens the spectrum of possibilities to conduct more research in this direction, as with this methodology more databases with fewer labeled data could be employed for the LV volume estimation and segmentation.

Declaration of Competing Interest

The authors declare that they have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors acknowledge financial support from the Conselleria d'Educació, Investigació, Cultura i Esport, Generalitat Valenciana (grants AEST/2019/037 and AEST/2020/029), 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), Spanish Ministerio de Ciencia, Innovación y Universidades (ref. CIIP-20192020).

References

[1] N. Townsend, L. Wilson, P. Bhatnagar, K. Wickramasinghe, M. Rayner, M. Nichols, Cardiovascular disease in Europe: epidemiological update 2016, Eur. Heart J. 37 (2016) 3232–3245, doi:10.1093/eurheartj/ehw334.

[2] E. Abdelmaguid, J. Huang, S. Kenchareddy, D. Singla, L. Wilke, M.H. Nguyen, I. Altintas, Left ventricle segmentation and volume estimation on cardiac MRI using deep learning, ArXiv Comput. Vis. Pattern Recognit. (2018) http://arxiv.org/abs/1809.06247. (accessed September 23, 2020).

[3] M. Perez-Pelegri, J.V. Monmeneu, M.P. Lopez-Lereu, S. Ruiz-Espana, I. Del-Canto, V. Bodi, D. Moratal, PSPU-net for automatic short axis cine MRI segmentation of left and right ventricles, in: 2020 IEEE 20th Int. Conf. Bioinforma. Bioeng, Institute of Electrical and Electronics Engineers (IEEE), 2020, pp. 1048–1053, doi:10.1109/bibe50027.2020.00177.

[4] R.P.K. Poudel, P. Lamata, G. Montana, Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation, in: Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), Springer Verlag, 2017, pp. 83–94, doi:10.1007/978-3-319-52280-7_8.

[5] Q. Tong, C. Li, W. Si, X. Liao, Y. Tong, Z. Yuan, P.A. Heng, RIANet: recurrent interleaved attention network for cardiac MRI segmentation, Comput. Biol. Med. 109 (2019) 290–302, doi:10.1016/j.compbiomed.2019.04.042.

[6] Q. Tao, W. Yan, Y. Wang, E.H.M. Paiman, D.P. Shamonin, P. Garg, S. Plein, L. Huang, L. Xia, M. Sramko, J. Tintera, A. de Roos, H.J. Lamb, R.J. van der Geest, Deep learning–based method for fully automatic quantification of left ventricle function from cine mr images: a multivendor, multicenter study, Radiology 290 (2019) 81–88, doi:10.1148/radiol.2018180513.

[7] C. Chen, W. Bai, R.H. Davies, A.N. Bhuva, C.H. Manisty, J.B. Augusto, J.C. Moon, N. Aung, A.M. Lee, M.M. Sanghvi, K. Fung, J.M. Paiva, S.E. Petersen, E. Lukaschuk, S.K. Piechnik, S. Neubauer, D. Rueckert, Improving the generalizability of convolutional neural network-based segmentation on CMR images, Front. Cardiovasc. Med. 7 (2020) 105, doi:10.3389/fcvm.2020.00105.

[8] N.M. deSouza, E. Achten, A. Alberich-Bayarri, F. Bamberg, R. Boellaard, O. Clément, L. Fournier, F. Gallagher, X. Golay, C.P. Heussel, E.F. Jackson, R. Manniesing, M.E. Mayerhofer, E. Neri, J. O'Connor, K.K. Oguz, A. Persson, M. Smits, E.J.R. van Beek, C.J. Zech, Validated imaging biomarkers as decision-making tools in clinical trials and routine practice: current status and recommendations from the EIBALL* subcommittee of the, Eur. Soc, Radiol. (ESR) Insights Imaging 10 (2019) 1–16, doi:10.1186/s13244-019-0764-0.

[9] J.H. Cole, R.P.K. Poudel, D. Tsagkrasoulis, M.W.A. Caan, C. Steves, T.D. Spector, G. Montana, Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker, Neuroimage 163 (2017) 115–124, doi:10.1016/j.neuroimage.2017.07.059.

[10] V. Gulshan, L. Peng, M. Coram, M.C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, R. Kim, R. Raman, P.C. Nelson, J.L. Mega, D.R. Webster, Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, JAMA - J. Am. Med. Assoc. 316 (2016) 2402–2410, doi:10.1001/jama.2016.17216.

[11] J.P. Vigueras-Guillen, J. Van Rooij, H.G. Lemij, K.A. Vermeer, L.J. Van Vliet, Convolutional neural network-based regression for biomarker estimation in corneal endothelium microscopy images, in: Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS, Institute of Electrical and Electronics Engineers Inc., 2019, pp. 876–881, doi:10.1109/EMBC.2019.8857201.

[12] A. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, Nature 542 (2017) 115–118, doi:10.1038/nature21056.

[13] G. Gonzalez Serrano, G.R. Washko, R. San José Estépar, Deep learning for biomarker regression: application to osteoporosis and emphysema on chest CT scans, in: E.D. Angelini, B.A. Landman (Eds.), Med. Imaging 2018 Image Process., SPIE, 2018, p. 52, doi:10.1117/12.2293455.

[14] G. González, G.R. Washko, R.S.J. Estépar, M. Cazorla, C.Cano Espinosa, Automated Agatston score computation in non-ECG gated CT scans using deep learning, in: E.D. Angelini, B.A. Landman (Eds.), Med. Imaging 2018 Image Process., SPIE, 2018, p. 91, doi:10.1117/12.2293681.

[15] C. Moreira, R. Sindhgatta, C. Ouyang, P. Bruza, A. Wichert, An investigation of interpretability techniques for deep learning in predictive process analytics, (2020). http://arxiv.org/abs/2002.09192 (accessed September 23, 2020).

[16] W. Samek, G. Montavon, S. Lapuschkin, C.J. Anders, K.-R. Müller, Toward interpretable machine learning: transparent deep neural networks and beyond, (2020). http://arxiv.org/abs/2003.07631 (accessed September 23, 2020).

[17] L. Chan, M.S. Hosseini, K.N. Plataniotis, A comprehensive analysis of weakly-supervised semantic segmentation in different image domains, Int. J. Comput. Vis. (2019) 1–24, doi:10.1007/s11263-020-01373-4.

[18] C. Cano-Espinosa, G. Gonzalez, G.R. Washko, M. Cazorla, R.S.J. Estepar, Biomarker localization from deep learning regression networks, IEEE Trans. Med. Imaging. 39 (2020) 2121–2132, doi:10.1109/TMI.2020.2965486.

[19] S. Wang, W. Chen, S.M. Xie, G. Azzari, D.B. Lobell, Weakly supervised deep learning for segmentation of remote sensing imagery, Remote Sens. 12 (2020) 207, doi:10.3390/rs12020207.

[20] Z. Huang, X. Wang, J. Wang, W. Liu, J. Wang, Weakly-supervised semantic segmentation network with deep seeded region growing, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Salt Lake City, 2018, pp. 7014–7023.

[21] G. Yang, C. Wang, J. Yang, Y. Chen, L. Tang, P. Shao, J.-L. Dillenseger, H. Shu, L. Luo, Weakly-supervised convolutional neural networks of renal tumor segmentation in abdominal CTA images, BMC Med. Imaging 20 (2020) 1–12, doi:10.1186/s12880-020-00435-w.

[22] C.-C. Hsu, K.-J. Hsu, C.-C. Tsai, Y.-Y. Lin, Y.-Y. Chuang, Weakly supervised instance segmentation using the bounding box tightness prior, Adv. Neural Inf. Process. Syst. 32 (2019) 6586–6597. (accessed September 23, 2020). https://github.com/chengchunhsu/WSIS_BBTP.

[23] A. Khoreva, R. Benenson, J. Hosang, M. Hein, B. Schiele, Simple does it: weakly supervised instance and semantic segmentation, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit, 2017, pp. 876–885.

[24] K.B. Girum, G. Créhange, R. Hussain, A. Lalande, Fast interactive medical image segmentation with weakly supervised deep learning method, Int. J. Comput. Assist. Radiol. Surg. 15 (2020) 1437–1444, doi:10.1007/s11548-020-02223-x.

[25] Z.H. Wang, Z.Z. Wang, Fully automated segmentation of the left ventricle in magnetic resonance images, (2020). http://arxiv.org/abs/2007.10665 (accessed May 20, 2021).

[26] Z.Z. Wang, Automatic localization and segmentation of the ventricles in magnetic resonance images, IEEE Trans. Circuits Syst. Video Technol. 31 (2021) 621–631, doi:10.1109/TCSVT.2020.2981530.

[27] G. Luo, G. Sun, K. Wang, S. Dong, H. Zhang, A novel left ventricular volumes prediction method based on deep learning network in cardiac MRI, in: 2016 Comput. Cardiol. Conf., Vancouver, BC, IEEE, 2016, pp. 89–92. https://ieeexplore.ieee.org/abstract/document/7868686.

[28] F. Zhu, Estimating left ventricular volume with ROI-based convolutional neural network, Turkish J. Electr. Eng. Comput. Sci. 26 (2018) 23–34, doi:10.3906/elk-1704-335.

[29] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), Springer Verlag,, 2015, pp. 234–241, doi:10.1007/978-3-319-24574-4_28.

[30] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-net: learning dense volumetric segmentation from sparse annotation, in: Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), Springer Verlag, 2016, pp. 424–432, doi:10.1007/978-3-319-46723-8_49.

[31] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, in: 32nd Int. Conf. Mach. Learn. ICML 2015, International Machine Learning Society (IMLS), 2015, pp. 448–456. https://arxiv.org/abs/1502.03167v3.

[32] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit, 2016, pp. 2921–2929. http://cnnlocalization.csail.mit.edu.

[33] P. Zhang, Y. Zhong, Y. Deng, X. Tang, X. Li, A survey on deep learning of small sample in biomedical image analysis, ArXiv Prepr. (2019) ArXiv1908.00473 http://arxiv.org/abs/1908.00473 . (accessed September 23, 2020).

[34] K.H. Zou, S.K. Warfield, A. Bharatha, C.M.C. Tempany, M.R. Kaus, S.J. Haker, W.M. Wells, F.A. Jolesz, R. Kikinis, Statistical validation of image segmentation quality based on a spatial overlap index, Acad. Radiol. 11 (2004) 178–189, doi:10.1016/S1076-6332(03)00671-8.

[35] Artifical Intelligence in EU medical device legislation, 2020. https://www.cocir.org/fileadmin/Position_Papers_2020/COCIR_Analysis_on_AI_in_medical_Device_Legislation_-_Sept._2020_-_Final_2.pdf (accessed November 9, 2020).