Cardiac Magnetic Resonance Phase Detection Using Neural Networks

Carles Garcia-Cabrera
School of Electronic Engineering and ML-Labs
Dublin City University
Dublin, Ireland
carles.garciacabrera6@mail.dcu.ie

Kathleen M. Curran
School of Medicine
University College Dublin
Dublin, Ireland
kathleen.curran@ucd.ie

Noel E. O'Connor
Insight SFI Research Centre for Data Analytics
Dublin City University
Dublin, Ireland
noel.oconnor@insight-centre.org

Kevin McGuinness
School of Electronic Engineering
Dublin City University
Dublin Ireland
kevin.mcguinness@dcu.ie

***Abstract*—The precision of cardiac magnetic resonance segmentation is an important area to investigate clinically and has received a lot of attention from the research community for its impact on the evaluation of cardiac functions. However, the correct identification of key time frames of cardiac sequences has received significantly less attention, especially in the MR domain, despite its great importance in the correct measurement of the Ejection Fraction, a key metric in diagnostics. In this paper, we present two deep learning regression methods to automate the otherwise time-consuming annotation process, with performance within the 1-2 frame distance error and almost instant calculation over short-axis images from a public dataset. Results are presented using publicly available data.**

***Index Terms*—Cardiac Imaging, Deep Learning, Phase Detection, Magnetic Resonance Imaging.**

I. Introduction

Cardiac Magnetic Resonance Imaging (MRI) image segmentation is an important step during quantitative analysis for the diagnosis of cardiovascular diseases. During the process, the image is partitioned into meaningful regions. The labelling of this information helps clinicians calculate and understand important features, such as the ejection fraction (EF), which are then used to determine whether subjects have a particular pathology or if surgery is necessary [1].

Electrocardiography (EKG or ECG) is a non-invasive diagnostic procedure employed to record the heart's electrical activity over time. This technique utilizes electrodes (usually between 12 and 15 [2]) placed strategically on a patient's skin to detect and measure the electrical signals generated by the heart's specialized cells during each cardiac cycle. The resulting data is represented as a graphical waveform, which provides invaluable insights into the heart's overall function, rhythm, and conduction system.

Precision during End-Diastolic (ED) and End-Systolic (ES) phase detection is a key aspect when measuring important features in cardiac functional analysis. Incorrect labels on these frames can lead to important errors in key clinical indicators [3], such as EF and global longitudinal strain (GLS), resulting in up to 10% error within the two and three frame difference [4], [5].

Developing techniques that suppress the need for accompanying cardiac MRI scan with ECG signals improves the robustness of algorithms that support cardiac analysis and prevent miscalculations or additional clinical efforts to label these phases manually.

Deep learning has been the golden standard for cardiac MRI segmentation in the past ten years. For this task, the Convolutional Neural Network (CNN) is the most important type of network, similar to other vision applications. In particular, U-shaped convolutional network variations have provided the best results [6], [7].

In this work, we target cardiac short-axis magnetic resonance images to automate the detection of ED and ES phases through deep learning in a regression task.

Our contributions are the following:

- **Report performance on public data:** the previous work on this topic reported results on private data [8], hence reproducing their results was impossible. We provide evidence of strong performance in an available open dataset M&Ms [9].
- **Comparison of two different architectures:** we studied the performance of different novel models, where we experiment with two elements that have shown great performance in a variety of problems that include sequences: (1) LSTM [10] and (2) Transformers [11].
- **Inference time:** the inference time of our models is 500 times faster than in previous work on the same problem [8]. While the results were obtained faster, the computational resources of the previous method are unknown.
- **Open source:** code is available at https://github.com/carlesgarciac/regression.

979-8-3503-6021-9/23/$31.00 ©2023 IEEE

II. Related Work

This section reviews the previous work related to our research and the two sequential encoders adopted for the regression task.

A. TempReg-Net

TempReg-Net [8], the previous work on cardiac MRI phase detection, used a deep learning approach that combined the Zeiler-Fergus model [12] encoder and a temporal decoder based on Recurrent Neural Networks (RNNs) with a particular loss function (Section III-C). The loss has two different components and penalizes frames labelled in the incorrect phase (either systolic or diastolic). The ground truth was synthesised to replicate the left ventricle volume changes in a standard cardiac cycle.

The results of [8] were obtained with private data from 420 patients.

In our work, we changed the first part of the architecture for the encoder part of a U-Net [13] due to its demonstrated performance in the segmentation task.

B. Sequential encoders

In our work, we adopted two of the most significant sequential modules of contemporary deep learning research. While both excel at processing sequences, there are some differences:

- LSTM: Long Short-Term Memory Networks (LSTMs) represent a class of RNNs widely adopted by deep learning-based models for their ability to model and capture long-range dependencies in sequential data. LSTMs address the vanishing gradient problem associated with traditional RNNs through the incorporation of specialized gating mechanisms, including input, forget, and output gates, which enable them to retain and update information over extended sequences selectively.
- Transformer: Transformers [11] are characterized by their attention-based mechanisms and parallelizable architecture. Unlike traditional RNNs and CNNs, Transformers rely on a self-attention mechanism, which enables them to capture intricate and long-range contextual information in input sequences, without being constrained by sequential processing. This parallelization of computation results in significantly reduced processing times.

III. Methodology

Our experiments were carried out on MRI sequences of public data, using two different novel deep learning architectures that we propose in this work. Details of our method and our experimentation are detailed in this section.

A. Data

In our experimentation, we employed the data released during Multi-Centre, Multi-Vendor and Multi-Disease Cardiac Image Segmentation Challenge (M&Ms). In particular, the dataset consists of 150 training cases, 34 for validation, and 136 for testing. For the pre-trained part of the network, we employed the M&Ms2 training data.

The subjects were scanned with scanners from four different vendors, two of them are present in the training set, and all of them are present in the testing one.

For our experimentation, we used a single middle slice for each time frame. An example of a short-axis mid-ventricle slice is depicted in Figure 1.

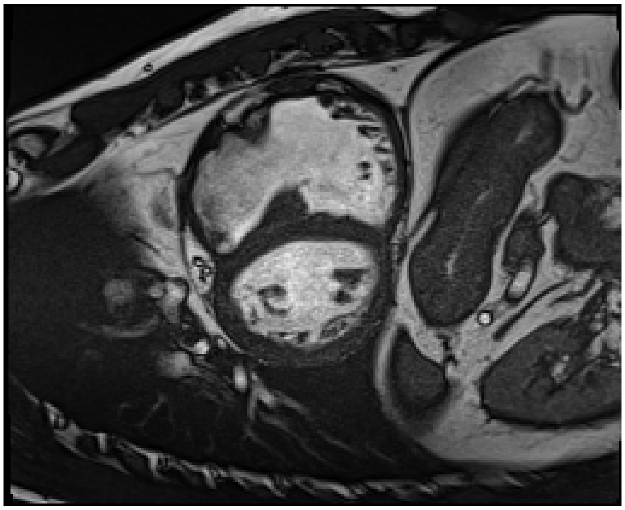

Fig. 1. Example of a mid-ventricle short-axis view from a CMR scan. Image from M&Ms2 [14].

B. Architecture

The model architecture comprises a CNN module (corresponding to the encoder part and the bottleneck of a U-net [13]) followed by a fully connected layer where CNN features are flattened. Then, the resulting features are passed to the part that extracts the features sequence-wise, an LSTM and a Transformer Encoder, in the first experiment and in the second experiment, respectively. The output of the sequence module is then connected to a second fully connected layer that outputs a vector corresponding to one element per frame in the sequence. The complete architecture is depicted in Figure 2.

The parameters chosen for our networks were the following:

- The CNN encoder was pre-trained in a segmentation task using the M&Ms2 data and then frozen while training the rest of the network. In particular, the parameters of the network were: 32 filters in the first out of five pairs of convolutional layers, and a max-pooling layer after each of the four first pairs of convolutional layers.
- The fully connected layers had 512 and one neuron in the input and output, respectively.
- The LSTM had the default PyTorch parameters, except that it was set to bidirectional. Both the input size and the hidden size were 512 neurons, and it had two layers.
- The Transformer Encoder was set to have 512 neurons as the input size, with four heads and two layers.

C. Experiment

In our experiment, we tested the performance of our two proposed networks, which were trained using a loss function consisting of two components:

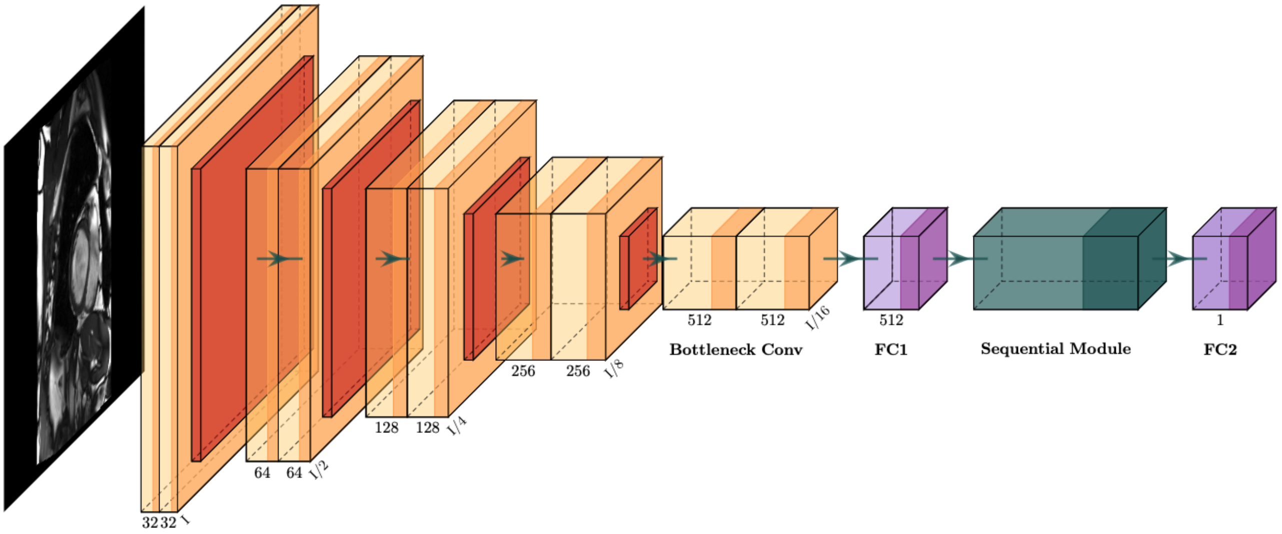

Fig. 2. The proposed network with the sequential module referring to an LSTM or a Transformer encoder in each experiment.

- The Mean Squared Error between the prediction and the synthetically generated signal (see equation 1).
- A temporal structured loss (see equation 2).

TempReg-Net [8] proposed this combination of losses.

$$y_k = \begin{cases} \left|\frac{k-N_{es}}{N_{es}-N_{ed}}\right|^{\delta}, & \text{if} \quad N_{ed} < k \leq N_{es} \\ \left|\frac{k-N_{es}}{N_{es}-N_{ed}}\right|^{\upsilon}, & \text{otherwise} \end{cases} \tag{1}$$

Where N is the ground truth for each phase, and is the time frame number. δ and υ are hyperparamaters set to 3 and $1/3$ respectively to mimic the behaviour of the left ventricle in the cardiac cycle.

$$\begin{aligned} L_{temp} &= \tfrac{1}{2}(L_{inc} + L_{dec}) \\ L_{inc} &= \tfrac{1}{T}\sum_{k=2}^{T} \mathbb{1}(y_k > y_{k-1}) \max(0, \eta_{k-1} - \eta_k) \\ L_{dec} &= \tfrac{1}{T}\sum_{k=2}^{T} \mathbb{1}(y_k < y_{k-1}) \max(0, \eta_k - \eta_{k-1}) \end{aligned} \tag{2}$$

η is the prediction.

To label the time frames, the maximum and the minimum of the signal are set as the ED and ES time frames. An example of the resulting signal is depicted in Figure 3. In this signal, the time frames corresponding to ED and ES are the first and eighth frames of the scan.

To evaluate the performance, we used the average Frame Difference (aFD) (see equation 3) to quantify the error.

$$aFD = \frac{1}{N}\sum_{t=1}^{N} |\hat{y}_t - y_t| \tag{3}$$

IV. Results

Table I details the performance in the test set of the M&Ms dataset for both architectures: aFD (less is better) and the detection time for each frame. We computed our method using an Nvidia GeForce RTX 2080Ti.

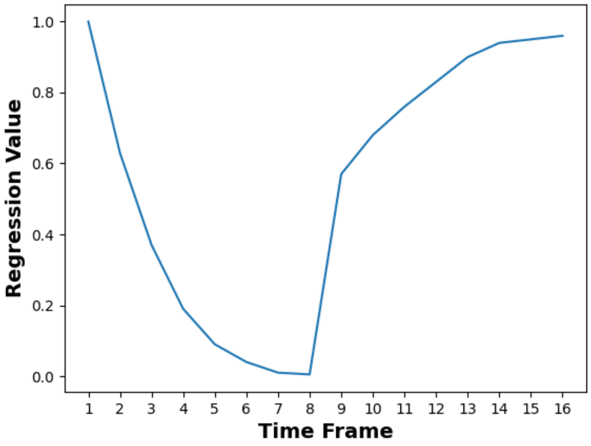

Fig. 3. An example of the output regression curve. The maximum (first frame) and the minimum (eighth frame) correspond to ED and ES time frames, respectively.

The results are promising since the range of our average frame difference remains below the 2 to 3 aFD, which can lead to important miscalculations for the EF. The LSTM performed marginally better than the transformer, while the detection time remained similar. Time-wise, we achieved almost instant calculations.

TABLE I
Time frame detection results: average frame difference and detection times.

Model	aFD ED	aFD ES	Detection Time (s)
CNN + LSTM	1.70	1.75	0.0028
CNN + Transformer	2.03	1.84	0.00246

V. Conclusions

Our models performed generally under the two-frame difference, even over data from unseen scanners. Our LSTM model outperformed the model with a transformer encoder. We hypothesize that the transformer encoder is performing with higher aFD due to the lack of fine-tuning of the parameters, which suggests there is potential for better results. Furthermore, the speed at which both models processed the data at testing time was much faster (0.0025 s) than others (more than one second).

We demonstrated that the performance of our proposed method is comparable to that of human annotators but at a much greater speed, easing the adoption of the algorithm in current clinical tools. Further training experimentation, such as training in conjunction with a segmentation task, improving training strategies, or joining the two proposed recurrent modules, proved to be promising research areas.

In future work, we speculate that the addition of optical flow [15] between frames could lead to improvements. As a direct input in a parallel encoder or indirectly by obtaining a feature from it. Moreover, incorporating more slices might provide additional benefits.

Acknowledgment

This work was funded by Science Foundation Ireland through the SFI Centre for Research Training in Machine Learning (18/CRT/6183).

References

[1] N. Kawel-Boehm, A. Maceira, E. R. Valsangiacomo-Buechel, J. Vogel-Claussen, E. B. Turkbey, R. Williams, S. Plein, M. Tee, J. Eng, and D. A. Bluemke, "Normal values for cardiovascular magnetic resonance in adults and children," *Journal of Cardiovascular Magnetic Resonance*, vol. 17, no. 1, p. 29, 2015.

[2] I. Vogiatzis, E. Koulouris, A. Ioannidis, E. Sdogkos, M. Pliatsika, P. Roditis, and M. Goumenakis, "The importance of the 15-lead versus 12-lead ecg recordings in the diagnosis and treatment of right ventricle and left ventricle posterior and lateral wall acute myocardial infarctions," *Acta Informatica Medica*, vol. 27, p. 35–39, 03 2019.

[3] B. H. Amundsen, "It is all about timing!," *JACC: Cardiovascular Imaging*, vol. 8, no. 2, p. 158–160, 2015.

[4] E. S. Lane, N. Azarmehr, J. Jevsikov, J. P. Howard, M. J. Shun-shin, G. D. Cole, D. P. Francis, and M. Zolgharni, "Multibeat echocardiographic phase detection using deep neural networks," *Computers in Biology and Medicine*, vol. 133, p. 104373, 2021.

[5] R. O. Mada, P. Lysyansky, A. M. Daraban, J. Duchenne, and J.-U. Voigt, "How to define end-diastole and end-systole?: Impact of timing on strain measurements," *JACC: Cardiovascular Imaging*, vol. 8, no. 2, pp. 148–157, 2015.

[6] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P.-A. Heng, I. Cetin, K. Lekadir, O. Camara, M. A. Gonzalez Ballester, G. Sanroma, S. Napel, S. Petersen, G. Tziritas, E. Grinias, M. Khened, V. A. Kollerathu, G. Krishnamurthi, M.-M. Rohé, X. Pennec, M. Sermesant, F. Isensee, P. Jäger, K. H. Maier-Hein, P. M. Full, I. Wolf, S. Engelhardt, C. F. Baumgartner, L. M. Koch, J. M. Wolterink, I. Išgum, Y. Jang, Y. Hong, J. Patravali, S. Jain, O. Humbert, and P.-M. Jodoin, "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.

[7] C. Chen, C. Qin, H. Qiu, G. Tarroni, J. Duan, W. Bai, and D. Rueckert, "Deep learning for cardiac image segmentation: A review," *Frontiers in Cardiovascular Medicine*, vol. 7, 2020.

[8] B. Kong and et al., "Recognizing end-diastole and end-systole frames via deep temporal regression network," in *LNIP, volume 9902*, (Cham), pp. 264–272, Springer International Publishing, 2016.

[9] V. M. Campello and et al., "Multi-centre, multi-vendor and multi-disease cardiac segmentation: The m&ms challenge," *MICCAI 2020*, vol. 40, no. 12, pp. 3543–3554, 2021.

[10] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, p. 1735–1780, nov 1997.

[11] A. Vaswani and et al., "Attention is all you need," 2017.

[12] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision – ECCV 2014* (D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, eds.), (Cham), pp. 818–833, Springer International Publishing, 2014.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," 2015.

[14] V. M. Campello, P. Gkontra, C. Izquierdo, C. Martín-Isla, A. Sojoudi, P. M. Full, K. Maier-Hein, Y. Zhang, Z. He, J. Ma, M. Parreño, A. Albiol, F. Kong, S. C. Shadden, J. C. Acero, V. Sundaresan, M. Saber, M. Elattar, H. Li, B. Menze, F. Khader, C. Haarburger, C. M. Scannell, M. Veta, A. Carscadden, K. Punithakumar, X. Liu, S. A. Tsaftaris, X. Huang, X. Yang, L. Li, X. Zhuang, D. Viladés, M. L. Descalzo, A. Guala, L. L. Mura, M. G. Friedrich, R. Garg, J. Lebel, F. Henriques, M. Karakas, E. Çavuş, S. E. Petersen, S. Escalera, S. Seguí, J. F. Rodríguez-Palomares, and K. Lekadir, "Multi-centre, multi-vendor and multi-disease cardiac segmentation: The m&ms challenge," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3543–3554, 2021.

[15] M. Zhai, X. Xiang, N. Lv, and X. Kong, "Optical flow and scene flow estimation: A survey," *Pattern Recognition*, vol. 114, p. 107861, 2021.