10.1148/ryai.2019180080

Appendix E1

Datasets

Training and Testing: Automated Cardiac Diagnosis Challenge (ACDC) Dataset

The ACDC dataset was made freely available by the University Hospital of Dijon (20). The dataset is divided evenly into five subgroups: healthy subjects and patients with myocardial infarction (ejection fraction of the left ventricle lower than 40% and several myocardial segments with abnormal contraction), dilated cardiomyopathy (diastolic left ventricular volume > 100 $mL/m^2$and an ejection fraction of the left ventricle lower than 40%), hypertrophic cardiomyopathy (left ventricular cardiac mass high than 110 g/m^2, several myocardial segments with a thickness higher than 15 mm in diastole and a normal ejection fraction), and abnormal right ventricle (volume of the right ventricular cavity higher than 110 mL/m^2 or ejection fraction of the right ventricle lower than 40%). The cine-MR images were acquired in breath hold with a steady-state free precession sequence in short axis orientation. Briefly, a series of cardiac short axis slices cover the left ventricle from the base to the apex, with a thickness of 5 mm. The spatial resolution ranges from 1.37 to 1.68 mm^2/pixel, and 28 to 40 frames per subject cover completely the cardiac cycle. Table E1 contains the patient demographics.

Validation on a Synthetic Dataset

This dataset enabled us to test the accuracy of the network on data acquired on different scanners with various field strengths from 1.5T to 3T and thus various image resolutions. Each simulation generates sequential cine frames with the corresponding ground-truth motion vectors that parameterize the motion from the first frame. However, we note that motion estimates are only generated for voxels within the myocardial tissue, and everywhere else are undefined. The output of the corresponding motion simulation is an $80 \times 80 \times 16 \times 3$ tensor, representing the required voxelwise deformation in the *x, y, z* directions. For each simulation, several parameters of the XCAT (eg, heart volume and shape) and MRXCAT (eg, signal-to-noise-ratio) were randomized.

Validation on a Real Dataset

The short axis cardiac MR images acquired from 33 subjects obtained using a GE Genesis Signal MR scanner and the FIESTA scan protocol were made freely available by the Hospital for Sick Children of Toronto, Canada. All the subjects were under the age of 19. Each patient's image sequence consisted of exactly 20 frames, and the number of slices along the long axis of the subjects ranged between 8 and 15. Spacing-between slices ranged between 6 and 13 mm. Each image slice consisted of 256×256 pixels with a pixel-spacing of 0.93–1.64 mm.

Architecture and Training

Two different learning rate schedules were compared: a constant learning rate of 1×10^{-4}, and a step decay schedule initialized at 1×10^{-4} and reduced by half every 10 epochs. We also hypothesized that due to the large variations in the heart volume of the subjects (eg, normal versus hypertrophy) as well as in the motion magnitude (eg, frame 1 to frame 2 versus frame 1 to

systole), the number of learnable parameters might limit the network performance. To test this, we compared our network with a similar architecture with fewer convolutional layers and fewer feature maps per layer (Fig E1). As shown in Figure E2, the combination of an architecture with more parameters and a step decay training schedule are closer to optimal. This combination was used for the rest of our experiments.

Evaluation Methods

We conducted a comprehensive evaluation of the performance of CarMEN by comparing its performance to that of a range of state-of-the-art methods we chose based on how widely used and accessible they were. These methods are summarized in Table E2. The hyperparameters of the Vampire method were selected as described by Gigengack *et al* (30) for cardiac applications, whereas the simplified Elastix package had a fully automated implementation that did not require tuning. For the other three methods, we obtained optimal parameters by using 50 subjects from the ACDC *training* dataset through a wide parameter sweep. The regularization parameters which maximized the average Dice similarity coefficient (DSC) on the training dataset were then used in our testing experiments.

The design and hyperparameters for the ITK BSpline methods were chosen initially from the default values as these provided reasonable results. Using 50 subjects form the ACDC *training* datasets, these parameters were then optimized to obtain the best Dice coefficient. However, we note that few parameters affected the Dice coefficient. This is likely due to the large variation in the patient population, ie, it is difficult to optimized for normal subjects and patients with hypertrophic cardiomyopathy simultaneously.

In summary, for the first method (BSpline1) we used a limited-memory Broyden-Fletcher-Goldfarb-Shanno (LBFGSB) optimizer with a gradient convergence tolerance equal to 1×10^{-5}, a maximum number of corrections and a maximum number of function evaluations set to 5 and 1000, and a cost function convergence factor set to 1×10^{7}. We optimized an image correlation loss function for 100 iterations with a linear interpolator and a meshsize equal to [8,8,8]. For the second method (BSpline2) we used as gradient descent optimizer with a learning rate set to 5, a convergence minimum value and window size set to 0.0001 and 5, and shrink factors and smoothing sigmas per level set to [6,2,1]. We optimized the image Mattes mutual information loss function for 100 iterations. The third method (BSplineElastix) did not required any hyperparameters as these are automatically estimated by the software. The values for the Vampire method were selected as described by Gigengack et al (30) as these parameters were optimized using an XCAT (30) synthetic cardiac dataset similar to ours. For the dDemons method, the fluid (SigmaFluid), diffusion (SigmaDiff), noise (Alpha) regularizers and parameters were sweep. The most relevant parameters are shown in Figure E3.

Results

We have provided additional figures of our results. Figure E4 is an outlier example in which CarMEN fails to accurately model the motion of two subjects with hypertrophic cardiomyopathy and abnormal right ventricle. Figure E5 shows the results from different slices (from base to apex) of a subject with hypertrophic cardiomyopathy. We have also included figures showing the dice similarity coefficient (Fig E6) and end-point-error (Fig E7) for the two cardiac phases used with the synthetic dataset separately.

References

36. Jan M. FAIR: Flexible Algorithms for Image Registration. SIAM, 2009. https://archive.siam.org/books/fa06/. Accessed July 1, 2018.

37. Kroon DJ, Slump CH. MRI modality transformation in demon registration. IEEE, 2009; 963–966.

Table E1: Patient Demographics

Subgroup	Average Height (cm)	Average Weight (kg)	Average BMI (m/kg^2)
Dilated LV Cardiomyopathy	171 ± 10	76 ± 17	26 ± 5
Hypertrophic LV Cardiomyopathy	170 ± 9	80 ± 20	28 ± 5
Myocardial Infarction	172 ± 6	74 ± 12	25 ± 5
Normal	172 ± 10	78 ± 14	26 ± 4
Abnormal Right Ventricle	170 ± 10	77 ± 21	27 ± 7

Table E2: Summary of Registration Software Tools and Methods Used

Name (author)	Method	Reference	Website Link
SimpleITK	BSpline1	(28,29)	http://simpleitk.readthedocs.io/en/master/Examples/ImageRegistrationMethodBSpline1/Documentation.html
SimpleITK	BSpline2	(28,29)	http://simpleitk.readthedocs.io/en/master/Examples/ImageRegistrationMethodBSpline2/Documentation.html
SimpleElastix	BSpline	(28,29)	https://github.com/SuperElastix/SimpleElastix
FAIR	Vampire	(36)	https://github.com/C4IR/FAIR.m
Kroon	dDemon	(37)	https://www.mathworks.com/matlabcentral/fileexchange/21451-multimodality-non-rigid-demon-algorithm-image-registration