Automatic Quality Control of Cardiac MRI

Predicting Segmentation Accuracy in Large-scale Population Imaging

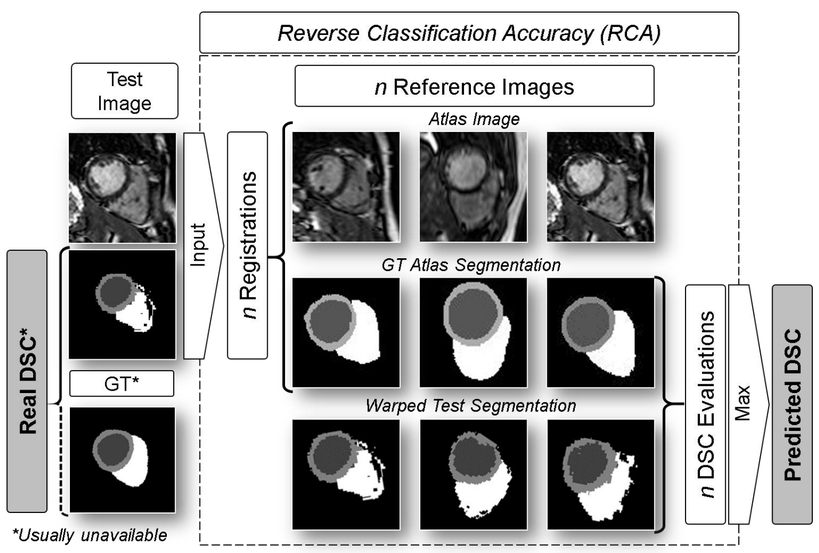
R. Robinson[1], V. Valindria[1], W. Bai[1], H. Suzuki[2], P. M. Matthews[2], C. Page[3], D. Rueckert[1], B. Glocker[1]

Motivation

The trend towards large-scale studies, poses new challenges for quality control (QC). This is a critical issue when automatic processing tools such as image segmentation methods are employed to derive quantitative measures for further analyses. It is important to be able to detect when an automatic method fails. As visual QC of individual segmentations is infeasible at large-scale, and ground truth is unavailable, we explore an automated approach for segmentation QC in population imaging.

Fig. 1: (Right) The RCA workflow employing single-atlas registration to find a prediction for segmentation quality in the absence of GT. (Below) Class labels for the cardiac MR segmentations.

Class	Region
1	LV Cavity
2	LV Myocardium
4	RV Cavity



1. Methods

We employ Reverse Classification Accuracy (RCA) [1] to predict the quality of automated 3D cardiac MRI segmentations. We validate RCA for application to cardiac data. In RCA the image and its automated segmentation are treated as an atlas with labels acting as *pseudo* ground truth (GT). Atlas registration and label-fusion are performed to a set of reference images with known GT. Similarity metrics, e.g. DSC [2], are calculated; the highest acting as a prediction of segmentation accuracy.

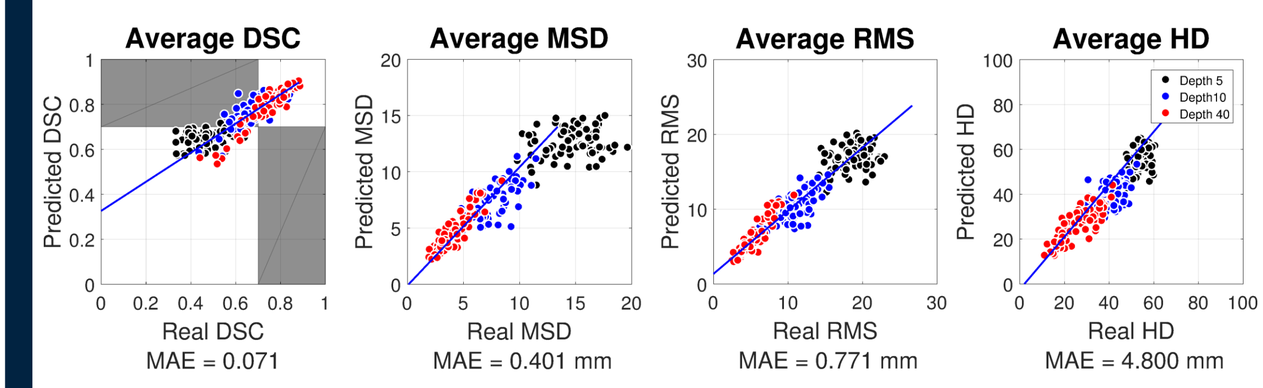


Fig. 2: Scatter plots demonstrating the correlation between predicted and real evaluation metrics for 240 automated segmentations when compared with their ground truth labels.

Class	Acc.	MAE DSC	MAE MSD (mm)	MAE RMS (mm)	MAE HD (mm)
1	0.883	0.054	0.276	1.001	10.691
2	1.000	0.033	0.242	0.754	6.917
4	0.813	0.080	0.482	1.179	6.482
All	**0.829**	**0.071**	0.401	0.771	4.800

Pred. DSC \ Real DSC	< 0.7	≥ 0.7	Total
< 0.7	94	0	94
≥ 0.7	41	105	146
Total	135	105	**240**

Fig 3: (Left) Table showing the mean absolute error (MAE) for each of the evaluation metrics across all 240 image-segmentations. RCA was able to correctly classify 83% of the segmentations as either 'poor', DSC $\in$ [0.0 0.7), or 'good', DSC $\in$ [0.7 1.0]. (Right) confusion matrix for the 2-class classification accuracy case.

2. Validation Results

- 240 automated segmentations. 20 reference images.
- RCA gave excellent per-case prediction rate of 83%
  - 'poor': DSC $\in$ [0.0 0.7)
  - 'good': DSC $\in$ [0.7 1.0]
- Good predictions or individual classes.
- Low Mean absolute error (MAE) across all metrics
- RCA valid for use on automated cardiac MR segmentations.

3. UK Biobank Results

- 7,425 automated cardiac segmentations from UK Biobank
- Manual QC scoring based on class 2: scores from 0 to 6 with -1 denoting a poor field of view.
- Good correlation between RCA predictions and manual QC scores with $p < 0.001$.
- Able to identify:
  - ✓ incorrect manual labels (see Fig. 4c)
  - ✓ poor segmentations in individual slices (Fig. 4b)
  - ✓ Good segmentations.

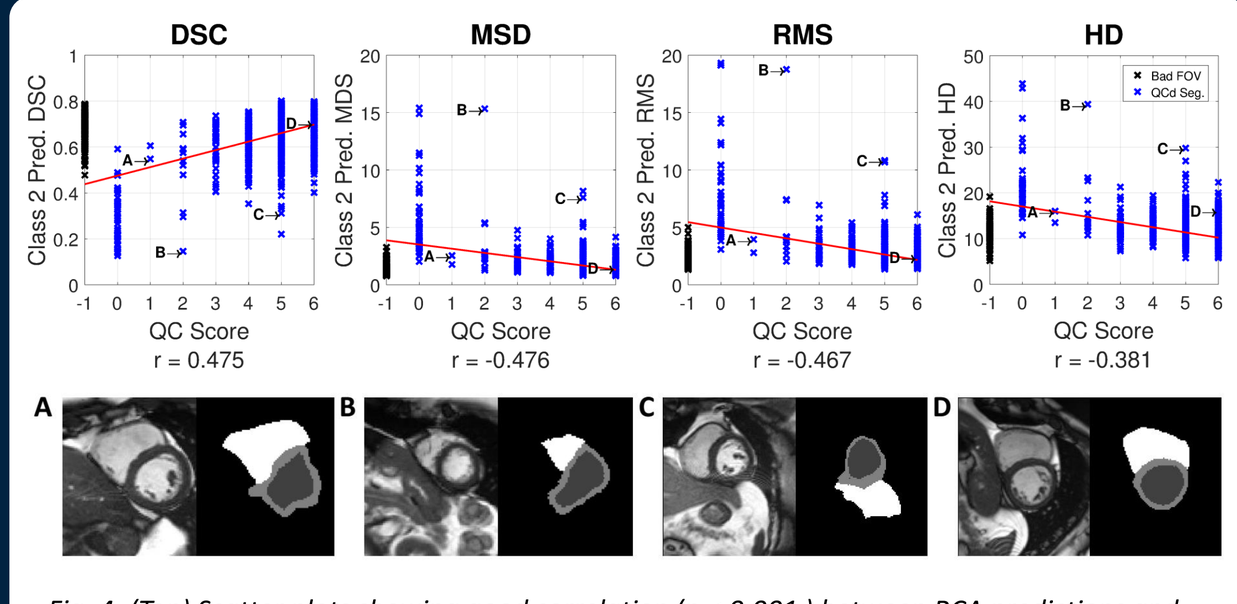


Fig. 4: (Top) Scatter plots showing good correlation ($p < 0.001$) between RCA predictions and manual QC scores on 7,425 automated segmentations of UK Biobank cardiac MRI. (Bottom) A-D examples of segmentations with manual score 1, 2, 5 and 6 respectively.

4. Conclusion

We show:

- Successful prediction of cardiac MR segmentation accuracy,
- Consistently good results across all evaluation metrics,
- Feasibility of employing RCA in large-scale studies.

References

[1] Valindria, V. V, Lavdas, I., Bai, W., Kamnitsas, K., Aboagye, E. O., Rockall, A. G., Rueckert, D., Glocker, B. (n.d.). "Reverse Classification Accuracy : Predicting Segmentation Performance in the Absence of Ground Truth", *TMI,* pp. 1–10, Feb 2017.

[2] L. R. Dice, "Measures of the Amount of Ecologic Association Between Species," *Ecology*, vol. 26, pp. 297-302, Jul 1945.

[1] BioMedIA Group, Department of Computing, Imperial College London, UK
[2] Division of Brain Sciences, Dept. of Medicine, Imperial College London, UK
[3] Clinical Innovation and Digital Platforms, GlaxoSmithKline Research and Development, Uxbridge, UK

Contact: r.robinson16@imperial.ac.uk