Exploring Pixel Value Scaling in the Application of Convolutional Neural Network U-Net Models for Segmentation of the Myocardium in Magnetic Resonance Images

Trygve Eftestøl[1], Mina Farmanbar[1], Tomas R Choat[1], Otto N Ljosdal[1], Mathias Dyvik[1], Casper Cappelen[1], Vidar Frøysa[2], Jørn Kværness[3], Gøran J Berg[1], Stein Ørn[2]

[1] Department of Electrical Engineering and Computer Science , Faculty of Science and Technology, University of Stavanger, Stavanger, Norway
[2] Stavanger University Hospital, Stavanger, Norway
[3] Good Clinical Research, Oppdal, Norway

Abstract

The aim of this work is the evaluation of deep learning models for automatic segmentation of the myocardium in Late Gadolinium Enhanced Cardiac Magnetic Resonance Images (LGE-CMRI). Two deep learning models trained on dataset A are applied to a separate dataset B. Due to variations in pixel value distributions between the two datasets, different scaling strategies for preprocessing the images before the model applications are explored. The deep learning models are evaluated: a Modified U-Net adding residual connections and a Multiresolution U-Net. The maximum pixel value, pv_m, over all slices for each patient specific image set is determined. To standardize pixel values in dataset (B), scaling is performed by dividing each by various values, based on both a fixed value and on the maximum pixel value. For each set of scaled cardiac MRI scans, the two models are applied. The highest Jaccard scores (median (25 percentile, 75 percentile)) on the test data for dataset A was 0.72 (0.60, 0.73) scaling with max-value scaling. For dataset B, the multiresolution model with floor-scaling, gave the highest Jaccard score of 0.65 (0.50, 0.70). This study highlights the significance of the dataset when assessing a model trained and tested on one dataset against another with varying pixel value distributions. The findings indicate that scaling could be effectively employed in semi-automated segmentation.

1. Introduction

Late gadolinium-enhanced cardiac magnetic resonance (LGE-CMR) is a common image modality for patients with myocardial infarction (MI). A gadolinium-based contrast dye is injected into the bloodstream of the patient. After a delay of approximately 10, the scarred tissue will appear bright and normal tissue will appear dark using a T1-weighted image sequence. The measurement of MI size is essential in assessing these patients, so it is important to delineate the myocardium.

Figure 1 presents four LGE-CMR images of a patient with myocardial infarction where a cardiologist has manually delineated the inner and outer walls of the left ventricle (the endo- and epicardium). Precise segmentation of the myocardium is crucial [1]. This is often a mandatory step prior to the analysis of such images for various purposes like relating arrhythmic heart rate to scar size [2] or transmurality [3], analysing myocardial infarction scar properties to asess myocardial injury [4, 5] and development of methods to detect microvasular obstruction in scarred regions [6]. Manual segmentation is time-consuming, and the use of CMRI is increasing. Therefore, the need for automatic segmentation is warranted [7, 8].

In this study, we applied two deep learning models, initially trained and evaluated on a dataset (A) during a student project, and then applied to a new dataset (B). Due to variations in pixel value distributions between the two datasets, we explored different scaling strategies for preprocessing the images before applying the models.

2. Materials and Methods

The LGE-CMR image data sets used in this study are referred to as dataset A and dataset B. The data were provided by the Department of Cardiology at Stavanger University Hospital. All images in both datasets are in a DICOM format with manual segmentations delineated by marking points along the myocardial walls.

Dataset A consists of 51 patients with myocardial infarction. This was further split into training (30), test (13), and validation (8) sets. The deep learning models for automatic segmentation were trained and evaluated using this

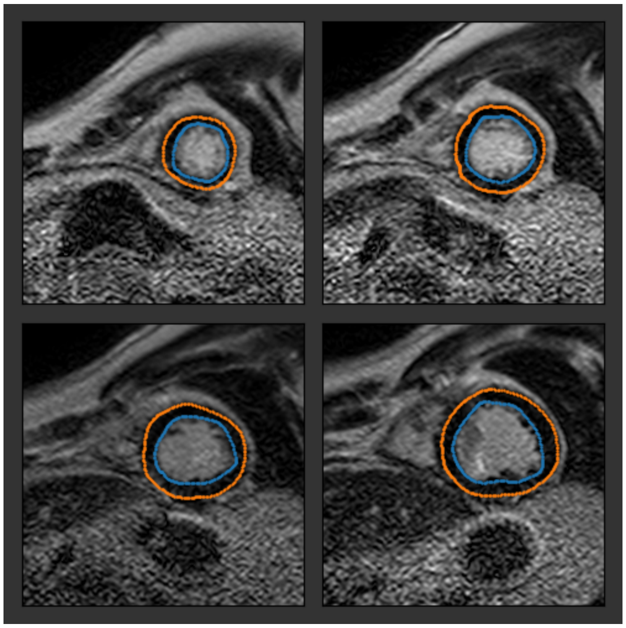

Figure 1: Original images with manually delineated endo- and epicardium.

dataset.

Dataset B was collected at a later time and consists of patients both with and without myocardial scars. Dataset B includes MRI scans from 58 patients and is used primarily for testing purposes.

Several deep neural network models were trained to segment the myocardium using variations of the U-Net architecture [9], commonly used for image segmentation. The U-Net model uses a symmetrical encoder and decoder. The encoder extracts features from the input image which are then fed into the decoder. In the decoder, an image is generated based on the features. An important aspect of U-Net is the direct connections between the encoder and decoder. This ensures the preservation of spatial information. In this study, we used models trained on dataset A during a student project: the Modified U-Net and the Multiresolution (MultiRes) UNet. Both models are buildt upon the U-Net architecture but incorporate more complex convolutional layers, aiming to enhance the robustness of segmentation results, particularly in the presence of outliers [10]. Each of these models differ in levels of complexity and computational requirements. In the student project, Standard U-Net, Modified U-Net and MultiRes U-Net models were trained using grid search on a number of parameters such as number of filters, batch size, dropout, learning rate. Serveral hundred combinations were explored.

The main performance results from the student project showed Jaccard scores of 0.576 ± 0.288 for the Standard U-Net, 0.731 ± 0.011 for the Modified U-Net and a score of 0.727 ± 0.013 for the MultiResU-Net.

During preprocessing, the images of set A were scaled by dividing each pixel value by 256. This was based on the assumption that 256 represents the maximum pixel value in this older dataset. This scaling process results in a normalization of all the pixel values into the range 0-1. All images are cropped to 256×256 pixels.

Dataset B is more recent, and the dynamic range varies between patients. Therefore a variety of scaling strategies are explored. For each patient, The maximum pixel value, pv_m, across all slices was determined.

To standardize the pixel values, pv, scaling was performed by dividing each by various values, based on both a fixed value and the maximum pixel. These scaling strategies include:

- Standard scaling corresponding to the one used on dataset A:

$$s_{std} = 256.$$

- Scaling by the maximum value in the image set:

$$s_{max} = pv_m.$$

- Scaling by a value corresponding to a lower number of bits:

$$s_{floor} = 2^{\lfloor log2(pv_m) \rfloor}.$$

- Scaling by a value corresponding to a higher number of bits:

$$s_{round} = 2^{\lfloor log2(pv_m)+1/2 \rfloor}.$$

- Scaling by a value corresponding to the nearest number of bits:

$$s_{ceil} = 2^{\lceil log2(pv_m) \rceil}.$$

The various scalings are thus achieved by scaling each pixelvalue. For example, the standard scaling defined as :

$$pv_{std} = pv/s_{std}$$

For each set of scaled images, the Modified U-Net and the MultiRes U-Net models are applied and their performance were evaluated using the Jaccard index. The jaccard index measures the degree of overlap between the manual segmentation mask, (M_m), and the predicted segmentation mask,(M_p). It is computed by determining the number of pixels, N_i, in the intersection between M_m and M_p. N_i is then divided by the number of pixels, N_u, in the union between M_m and M_p. Thus,

$$Jaccard = N_i/N_u.$$

The values are presented as the median (25 percentile, 75 percentile).

The experiments are conducted on both dataset A and dataset B for comparison.

3. Results

The results of the two models using the scaled test data as input are shown in Table 1 for dataset A and B. and B, respectively.

For dataset A, the Modifies U-Net with maximum value scaling has the highest performance, while the MultiRes U-Net with floored scaling performs best with dataset B. Note the drop in performance when comparing these results for the two datasets.

The interquartile ranges suggests large variations in the performance. We look at some examples from dataset B to get an impression of the variations in the segmentation quality.

The MultiRes U-Net model with s_{floor} scaling achieved the highest median performance. Figure 2 presents a sequence of slices for a given patient, applying the best-performing model with floor scaling with Jaccard scores ranging from 0.712 to 0.849.

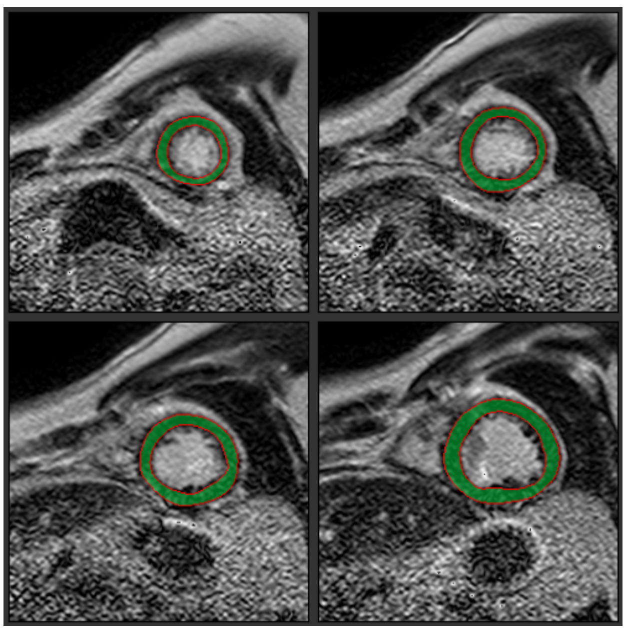

Figure 2: Original images with predicted masks overlaid for all images from one patient with a median Jaccard score of in the range 0.712-0.849.

Figure 3 shows a sequence of slices for a given patient, with the same model and scaling as above ranging from 0.229 to 0.529.

Figure 4 shows a sequence of slices for a given patient, with the same model but standard scaling. Note the differences in the segmentation results.

Interestingly, for dataset A, the Modified U-Net model was better for all scalings except the standard scaling, with the median value of the Jacard score typically 2-4 points

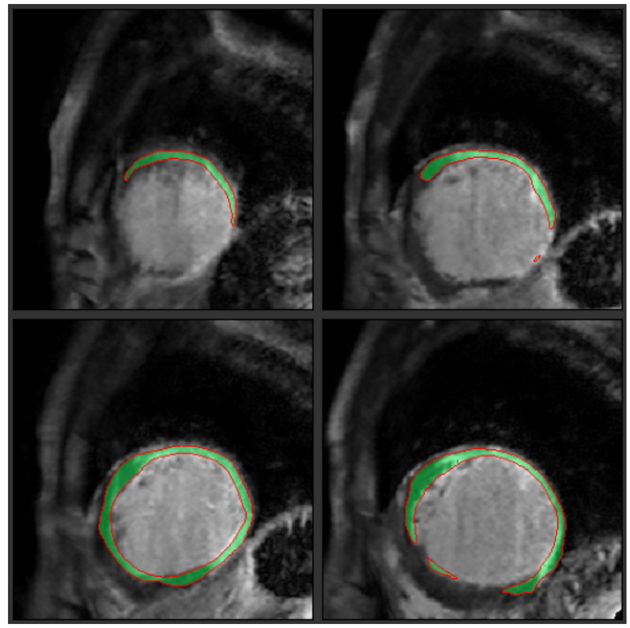

Figure 3: Original images with predicted masks overlaid for all images from one patient with a median Jaccard score of in the range 0.229-0.529.

higher than for the Multoiresolution model. For dataset B, the MultiResU-Net model was better with scores typically 3-4 points higher than for the other model. It is also worth noting that the interquartile ranges for the Jaccard scores are larger for dataset B than for dataset B. One might speculate that dataset B contains images that are too different from those used in the training of the model.

The interquartile ranges of the scores for the best performing model indicates that the quality of the segmentations vary substantially. This is illustrated by the figures which have been chosen to show both low and high-quality segmentations.

In the example with the best segmentations, similar visualisations showed mininal changes with variations in scaling. For the examples with the poorly performing segmentations we see clear differences between the scaling methods. These images also appear darker. One might speculate that further experiments with other scaling methods might improve the performance further.

Our results seems to be well aligned with Chen et al. who reported similar performance metrics in a larger dataset, using a U-Net architecture in their experiments [11].

4. Conclusion

We have demonstrated a deep neural network solution for automatically segmentting of LGE-CMR images. The drop in performance was notable when compared to the

Model architecture	s_{std}	s_{round}	s_{floor}	s_{ceil}	s_{max}
Modified U-Net (A)	0.70(0.60,0.73)	0.70(0.60,0.73)	0.71(0.58,0.73)	0.70(0.60,0.73)	**0.72(0.60,0.73)**
MultiRes U-Net (A)	0.70(0.60,0.71)	0.66(0.59,0.71)	0.67(0.60,0.71)	0.66(0.59,0.71)	0.70(0.64,0.71)
Modified U-Net (B)	0.59(0.43,0.68)	0.59(0.40,0.67)	0.61(0.44,0.69)	0.53(0.35,0.68)	0.59(0.40,0.68)
MultiRes U-Net (B)	0.62(0.52,0.71)	0.62(0.52,0.70)	**0.65(0.50,0.70)**	0.60(0.44,0.70)	0.64(0.52,0.70)

Table 1: Metrics for the two models with scaling.

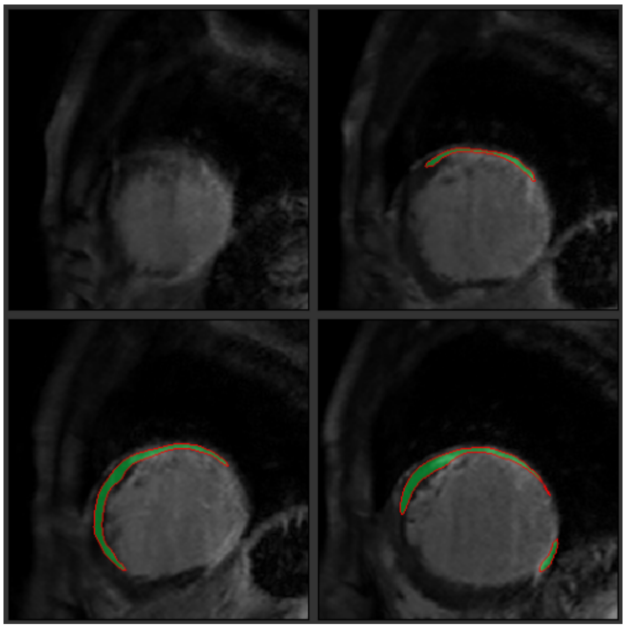

Figure 4: Original images with predicted masks overlaid for all images from one patient with a median Jaccard score of in the range 0.229-0.529.

performance on a test set from the dataset used to train the models. However, the results are encouraging, and one can use these models in a semi-automatic segmentation tool.

References

[1] Khened M, Kollerathu VA, Krishnamurthi G. Fully convolutional multi-scale residual densenets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. Medical Image Analysis 2019;51:21–45.

[2] Woie L, Eftestøl T, Engan K, Kvaløy JT, Nilsen DW, Ørn S. The heart rate of ventricular tachycardia following an old myocardial infarction is inversely related to the size of scarring. Europace 2011;13(6):864–868.

[3] Woie L, Engan K, Eftestøl T, Kvaløy JT, Ørn S. The relationship between transmurality of ischemic scars and the heart rate of ventricular tachycardia. Scandinavian Cardiovascular Journal 2015;49(5):241–248.

[4] Frøysa V, Berg GJ, Eftestøl T, Woie L, Ørn S. Texture-based probability mapping for automatic scar assessment in late gadolinium-enhanced cardiovascular magnetic resonance images. European Journal of Radiology Open 2021; 8:100387.

[5] Frøysa V, Berg GJ, Singsaas E, Eftestol T, Woie L, Orn S. Texture-based probability mapping in cardiac lge images: A novel automatic assessment of myocardial injury following revascularized stemi. medRxiv 2023;2023–09.

[6] Eftestøl T, Singsaas E, Engan K, Woie L, Ørn S. Automatic detection of microvascular obstruction in patients with myocardial infarction. In Computing in Cardiology Conference (CinC), 2015. IEEE, 2015; 757–760.

[7] Kim YC, Kim KR, Choi K, Kim M, Chung Y, Choe YH. Evcmr: a tool for the quantitative evaluation and visualization of cardiac mri data. Computers in Biology and Medicine 2019;111:103334.

[8] Engan K, Naranjo V, Eftesøl T, Ørn S, Woie L. Automatic segmentation of the epicardium in late gadolinium enhanced cardiac mr images. In Computing in Cardiology Conference (CinC) 2013. 2013; 631–634.

[9] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, part III 18. Springer, 2015; 234–241.

[10] Ibtehaz N, Rahman MS. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. Neural Networks 2020;121:74–87.

[11] Chen Z, Lalande A, Salomon M, Decourselle T, Pommier T, Qayyum A, Shi J, Perrot G, Couturier R. Automatic deep learning-based myocardial infarction segmentation from delayed enhancement mri. Computerized Medical Imaging and Graphics 2022;95:102014.

Address for correspondence:

Trygve Eftestøl
University of Stavanger, N-4036 Stavanger, Norway
tel./fax: +47-5183-2035/1750
trygve.eftestol@uis.no