Deep Learning Based Myocardial Segmentation for Free-Breathing Real-Time Cine MRI: Augmentation from Breath-Hold Scans

Hui Xue
hxue@stanford.edu

Abstract

MR Cine imaging is the gold-standard to visualize cardiac motion abnormality. While conventional cine MRI relies on lengthy breath-hold scans, free-breathing real-time cine MRI improves imaging efficiency and robustness. To effectively analyze real-time cine images, this study proposed a deep learning based solution to delineate myocardium. A neural net was developed by enhancing the U-net architecture with Resnet modules. Data augmentation and pre-processing workflow was introduced to utilize breath-hold cine training samples for real-time scans. Total training set consists of 44,428 images for endo and epi-myodardium, from N=1,308 patients. Good segmentation accuracy was measured on a test set of N=45 patients, with Dice ratio being higher than 0.9. This solution was integrated to clinical MR scanner and running in collaborative hospitals

Introduction

Cardiac motion abnormality is a strong biomarker for Cardiac vascular diseases (CAD). Current gold-standard to evaluate cardiac motion is the magnetic resonance imaging (Figure 1). The state-of-art method is to image the heart at multiple phases (time points in a heart beat). If playing all these images (e.g. often 30 images per heart beat) as a movie, it is called "Cine" imaging.

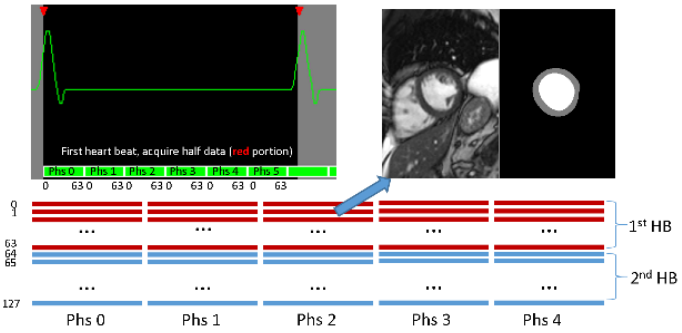

Figure 1. An illustration of breath-hold cardiac cine MR imaging. In this scheme, to acquire multiple images coving different phases of heart motion, patients are required to hold their breath. In this example, half of the data is acquired during the first heart beat and second half is acquired during the second heart beats. These two portions are assembled together to get the image. One image is shown here, together with manually segmented left ventricular. If there were any inconsistency between two heart beats, e.g. breathing motion, the image quality will degrade.

There are two major limitation of current method. First, the imaging requires patient to **hold the breath**, since the MR scan is performed by acquire a portion of data per heart beat and combine 3-7 heart beats to fill the data space (so-called "k-space" in MRI). There cannot be respiratory motion between heart beats; otherwise, the image will be contaminated by blurring. Second, the motion abnormality is often read out by expert watching the "Cine" movie (literally!). Resulting report is often a coarse answer as "Yes, with abnormality" or "No, free of abnormality". There is no quantitative reporting in the process. This leads to subjective, time-consuming image reading session, which takes bulk time from cardiologists' busy schedule.

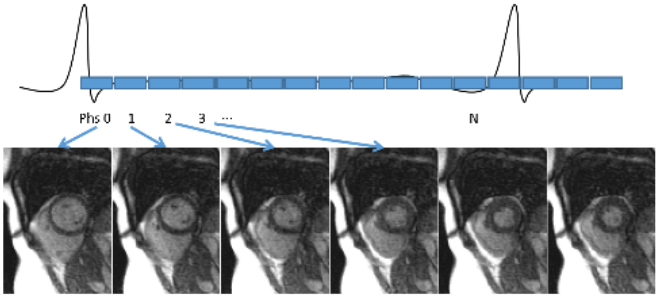

Figure 2. An illustration of free-breathing real-time cine imaging. In this scheme, all data required to assemble an image is acquired at once without splitting data acquisition into multiple heart beats. This improved robustness of imaging and also allow patients to breath freely. However, the price to pay is the lower image quality, as compared to Figure 1. But clinicians can still prefer real-time cine images if an automated analysis can be provided to them.

Recent improvement in MR imaging leads to a new scheme for data acquisition, called real-time cine imaging (**RT Cine**) [1]. As shown in Figure 2, instead of combining data from multiple heart beat, real-time acquisition will start at the R-wave trigger and continuously read out every MR image per 30-40ms. The advantage is no breath-holding required and imaging can be performed with very high efficiency. Because for the high imaging speed, the real-time cine imaging (Figure 2), compared to breath-hold scans (Figure 1), has inferior image quality, in terms of spatial resolution and signal-to-noise ratio. However, clinical practice showed cardiologists are tolerating lower image quality and start to embrace real-time cine for its efficiency and robustness.

The major barrier for cardiologists to fully adopt real-time cine imaging is the lack of automated analysis. Ideally, if myocardium can be automatically segmented, cardiac motion can be extracted as e.g. strain map, and reported to clinicians. Towards fully automated evaluation of cardiac motion with free breathing, this project aims to develop a deep learning based solution to automatically segment myocardium from real-time cine imaging.

Related work

While research had been conducted to segment breath-hold cine imaging using deep NN, no prior study is reported for free-breathing, real-time cine. The first published paper [2] for breath-hold cine segmentation utilized the fully convolutional network (FCN) [3]. This network was extended by adding encoding-decoding structures and expanding the filter depth while reducing the spatial resolution [4]–[7]. This led to major improvement in performance, as demonstrated by U-net [6] or Seg-net [5]. The U-net architecture was further extended by using recurrent [8] and residual modules [9]. More recent variation included V-net [4] and TernausNet [10], which kept the skip connections, but had different number of convolution filters and kernel size in each down/upsampling layers. Recent publications started to apply NNs for other cardiac imaging applications, such as automated analysis of T1 mapping image [11], cardiac flow images [12] and perfusion imaging [13].

Dataset

Initial datasets do not include labelled RT cine samples. To facilitate the task of real-time cine segmentation, a strategy was developed to start with breath-hold cine and gradually migrate to RT cine.

Imaging Both breath-hold imaging and real-time cine was performed at Barts Heart Centre, London, UK. Typical breath-hold imaging protocols are: field of view 360×270, imaging matrix 256×144, number of heart beats 7-9, 2x acceleration. A total of 9 to 12 slices are acquired to cover the heart. Reconstructed image resolution for breath-hold cine is then ~1.4mm^2. The rea-time cine has lower spatial resolution and faster acquisition, leading to lower signal-to-noise ratio: field of view 360×270, imaging matrix 160×120, 4x acceleration. Reconstructed image resolution for RT cine is thus ~2.25mm^2. Data was acquired with the required ethical and/or audit secondary use approvals or guidelines (as per each center) that permitted retrospective analysis of anonymized data for the purpose of technical development and protocol optimization and quality control. All data was anonymized and de-linked for analysis by NIH with approval by the NIH Office of Human Subjects Research OHSR.

Training Data Two group of training data was collected. a) Breath-hold cine images from N=1,208 patients were retrospectively collected for the training. For every case, the end-diastolic (ED, when the heart is at the most relaxation) and end-systolic phases (ES, when heart is at the most contraction) were selected. For all ED phases, the inner (endo) and outer (epi) boundary of myocardium was traced by clinicians. For all ES phases, only the endo boundary was traced. These give 28,084 endo and 14,042 epi images. b) After training an initial model using only the augmented breath-hold cine data, the preliminary model was applied to N=98 patients for 1,151 slices. Imaging were conducted between 20190412 and 20190515. For every case, both ED and ES phases were segmented for both endo and epi boundaries. The initial segmentation was checked for accuracy and corrected manually if required. This RT cine dataset was added for the final training.

Testing Data Real-time cine imaging data was retrospectively collected from N=45 patients, which constitutes 570 slices. The ED and ES phases were manually selected, and both endo and epi-myocardial boundaries were delineated. The trained model was applied to all test dataset. The NNs derived endo and epi mask was compared to the manual ground-truth.

Methods

NN architecture The U-net semantic segmentation architecture [6], [14] was modified for the cine myocardium detection. As shown in Figure 3, a ResNet module [15] was used as the building block. To simplify the neural net specification, one module consists of two convolution (CONV) blocks with same number of output filters. As shown in Figure 3, the downsampling and upsampling layers were connected with "Skip-connections". The input image went

through the input and then downsampling layers. While the spatial resolution was reduced, the number of filters increased. The upsampling layers increased the spatial resolution and reduced the number of filters. Final CONV layer outputted scores for background and foreground (LV mask). The output of NN was inputted into pixel-wise sigmoid to get the probability maps. All CONV layers used 3x3 with stride 1 and padding 1. More blocks or layers can be added to make NNs deeper.

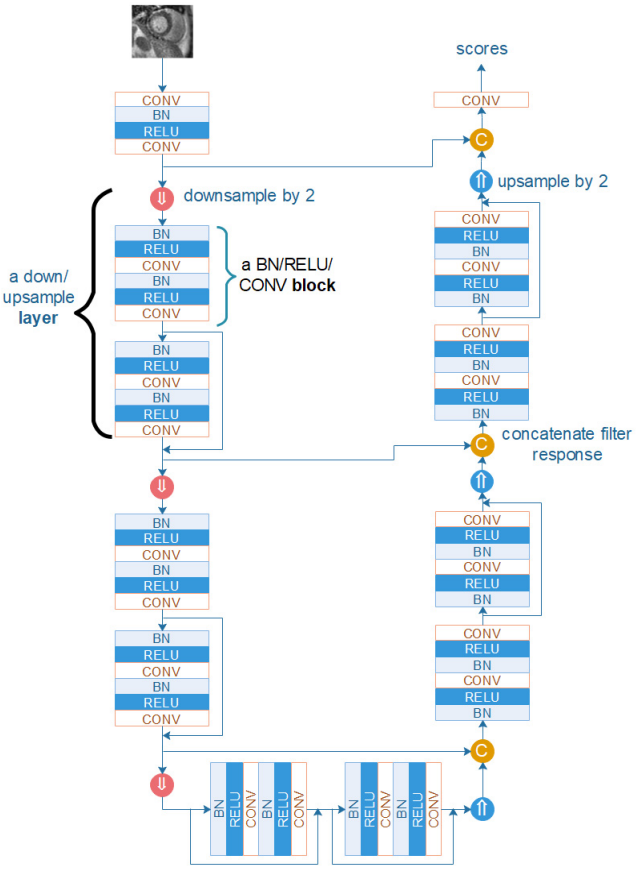

Figure 3. The schematic plot of neural nets trained in this study.

Data preprocessing As shown in Figure 4, training images were first resampled to 1mm^2 spatial resolution. The center portion of image (216×216mm^2) was selected, resulting in squared image of 216×216 pixels. The same process was applied to the labelled endo/epi masks.

Due to the gradually decreased receiver gain deeper into patient body, the cine images had intensity attenuation (Figure 5). Since this intensity variation can influence the NNs performance and may be difficult to simulate during data augmentation, a correction algorithm was developed. The time series of cine images was reformatted to *number_of_pixels* × *number_of_phases* data matrix. The PCA was performed along the time dimension. The first eigen mode (so-called eigen image) is a good representation of receiver gain variation and had the least amount of anatomical information. This eigen-image was spatially smoothed and fitted to low-order smoothing function as an estimation receiver gain. The corrected images were computed by dividing the estimated receiver gain. Figure 5 shows original RT cine image, estimated receiver gain and corrected image.

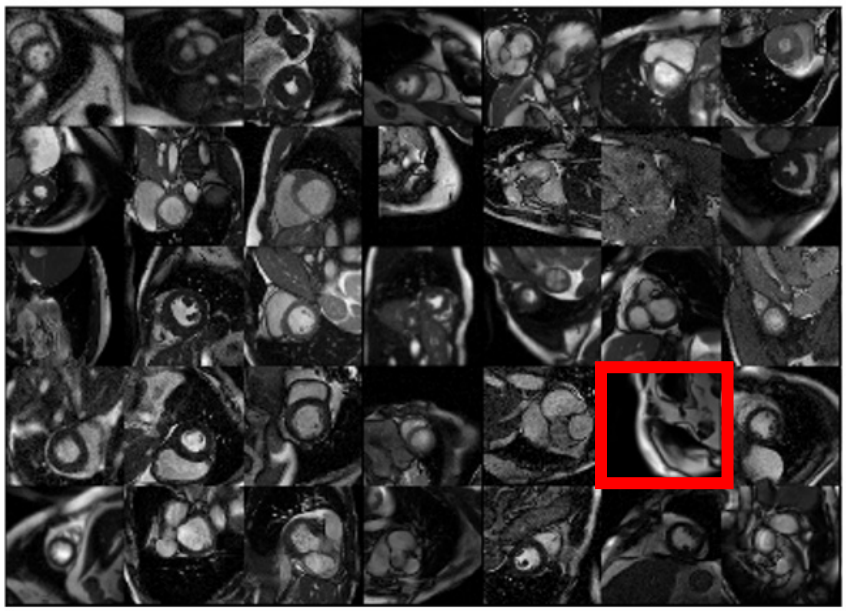

Figure 4. Training data after data augmentation. Every image here has the size of 216x216. Different amount of noise and blurring were added to images. Anatomical variation among samples are noticeable. Some images do not include myocardium (e.g. the one in red box), which should be reported as such by NNs.

Data augmentation and training Given different characteristics between breath-hold and real-time cine imaging and the fact that all available training data are from the former group, a strategy was developed for NN training. To mimic lower spatial resolution and SNR of RT cine, data augmentation was applied, including adding random noise and random

blurring. The noise level was randomly decided to be [0.05, 0.075, 0.1, 0.15, 0.2] times mean image intensities. The Gaussian blurring was applied with sigma being randomly selected as [0.5, 0.75, 1.0, 1.25, 1.5, 2.0, 2.5, 3.0] mm. Proposed NNs was implemented using PyTorch. The data augmentation was implemented as data transformation objects applied before extracting every mini-batch. Two NN models were trained for either endo and epi segmentation. All training was performed on a Microsoft Azure cloud node equipped with GPUs. The ADAM optimization was used with initial learning rate being 0.001. The betas are 0.9 and 0.999. Epsilon was 1e-8. Learning rate was reduced by x2 for every 10 epochs. Training took maximal 40 epochs and best model was selected as the one giving best performance on Dev set.

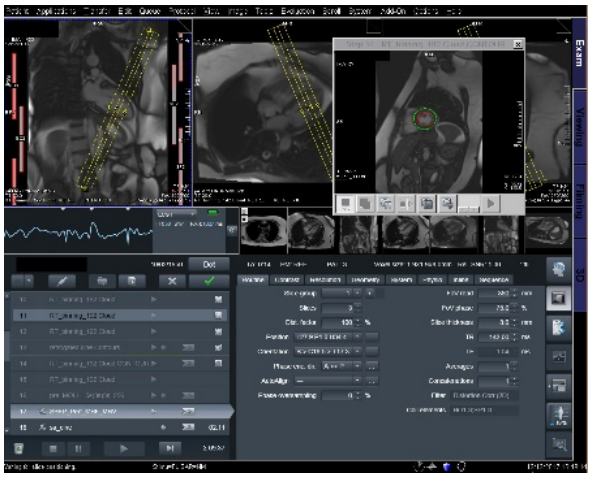

Figure 6. Integration of NN model on MR scanner for cine segmentation.

Evaluation and training data curation For every imaging slice of tested RT cine, the ED and ES phases were selected. The decision was made first whether myocardium was imaged in this slice (as shown in Figure 4). If imaging slice did not cover myocardium, the ground-truth label was left blank. Otherwise, the endo and epi boundaries were manually delineated. The Dice ratio was computed between automated segmentation and manual labelling. To measure regional segmentation error, the myocardial boundary distance (MBD) was computed as the mean distance between auto and manual endo/epi contours. The labelled RT cine data was added back for second round NN training before model deployment.

MR scanner integration The final model after adding the RT cine training samples was saved for scanner integration. An open-source framework (Gadgetron [14]) was used to load and apply the Pytorch model in MR scanners. The RT cine imaging protocol was modified to call up the NNs and generate myocardium contours on the fly. As shown in Figure 6, the results of AI models were sent back to scanner and presented to operators. This workflow was completely automated and in-line.

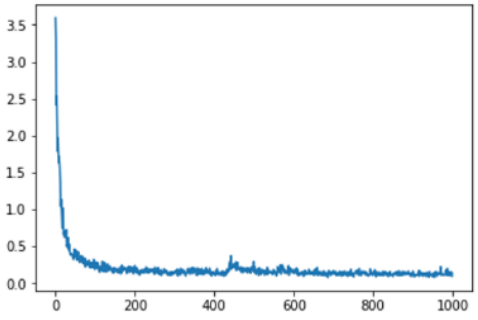

Figure 7. Cost vs. iteration plot.

Software developed for RT cine segmentation is shared at https://github.com/xueh2/cmr_cine.git

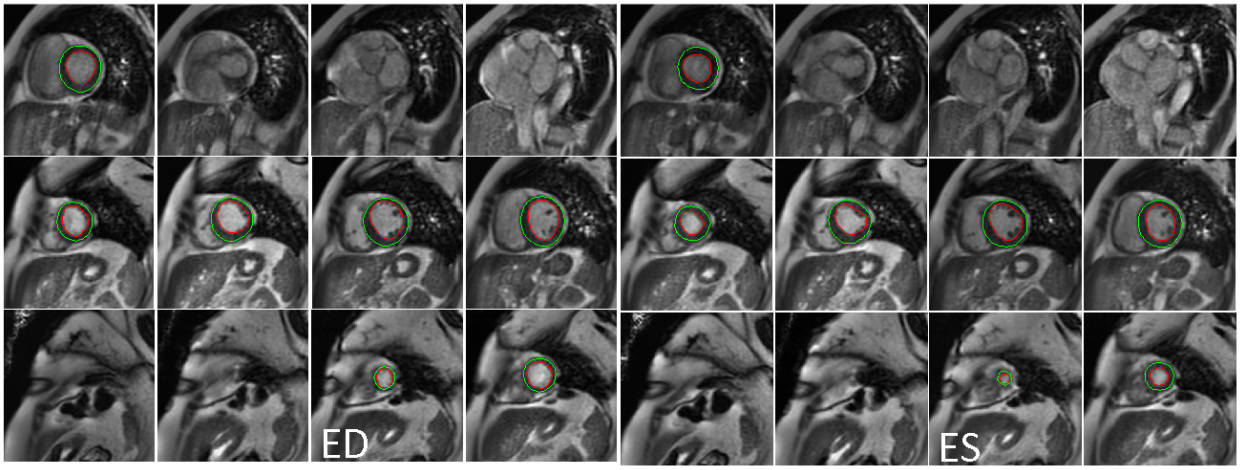

Figure 8. An example of NNs based segmentation on RT cine. The end-diastolic and end-systolic phases are shown. Given 12 slices were acquired in this scan, NNs were able to segment myocardium from very basal to apical planes. For those slices where imaging did not cover myocardium, NNs gave low probability.

Results

Proposed NNs and training strategy was applied to RT cine images. Data augmentation effectively mitigated the differences in SNR and resolution between training and testing datasets. Figure 7 gave the cost – iteration plot for endo training, while similar pattern was observed for epi models. The Adam optimizer effectively brought down the cost and allowed NNs to learn myocardial anatomy. Figure 8 gave RT cine segmentation for all slices for a patient. The trained NNs successfully segmented endo- and epicardium at all slices. Furthermore, for imaging slices which did not cover the myocardium, the NNs were able to correctly produce low probabilities, leading to "no segmentation".

For the test datasets of 570 RT cine slices, a total of 136 slices did not include the myocardium. Proposed NN solution correctly identified 129 slices by giving "no segmentation". For the slices where myocardium was imaged, the Dice ratio for endo and epi-myocardium was 0.95±0.07 and 0.96±0.06 for ED phase, and 0.93±0.10 and 0.96±0.05 for ES phase. The myocardial boundary distance was 0.31±0.72mm and 0.30±0.79mm for ED endo and epi boundaries. At the ES phase, boundary distance was 0.37±0.75mm and 0.33±0.51mm. Given the resampled image resolution of $1mm^2$, the mean myocardial boundary error was around 1 pixel.

Discussion

This study proposed a deep learning based solution to segment myocardium from free-breathing real-time cine MRI. Training process started from the conventional breath-hold cine data and migrated to real-time cine through data augmentation and bootstrap training data collection. Testing on independent RT cine datasets showed good segmentation was achieved for myocardial imaging slices with mean boundary error being ~1pixel. Further improvement is required to detect "negative" samples where myocardium was not included in FOV. Figure 9 gave an example of incorrect detection on a "negative" slice. More training samples covering broader imaging slice configuration can be necessary to improve detection rate of "negative" slices.

Figure 9. An example of incorrect NN detection. This slice was imaged above the myocardium, while the NNs gave “positive” detection. Possible solution could be to add these negative samples into training set.

One major barrier for developing medical imaging applications using deep learning is the cost to establish training set, in both expert hours and financial aspects. Computer vision community generates gigantic datasets of labelled images by hiring "common" workers through internet (e.g. "crowdsourcing" in Amazon Market Place). However, this may not be a feasible option for medical imaging data labelling, due to required professional knowledge and expertise. This study utilized the bootstrap training data collection to speedup curation of RT cine samples. Recent development of synthetic data labelling can learn a generative model from a small training set and synthesize more training data by sampling the learned probability distribution. This strategy is promising with the need of expert supervision.

Conclusion and Future work

This study proposed a deep learning based solution to delineate myocardium from real-time cine MRI. A neural net was implemented by combining the U-net architecture and Resnet modules. The trained NNs were integrated to MR scanner, enabling fully automated inline detection of myocardium. A boot strap training sample collection strategy was utilized, together with data augmentation, to achieve good segmentation performance. Further improvement could involve adding more "negative" samples to avoid false-positive detection.

Contribution

HX conceived the study, collected RT cine datasets, developed the neural net solution, implemented data preprocessing, conducted training and compiled results.

Reference

[1] J. D. Biglands, A. Radjenovic, and J. P. Ridgway, "Cardiovascular magnetic resonance physics for clinicians: part II," *J. Cardiovasc. Magn. Reson.*, vol. 14, no. 1, p. 66, 2012.

[2] W. Bai *et al.*, "Automated cardiovascular magnetic resonance image analysis with fully convolutional networks," *J. Cardiovasc. Magn. Reson.*, vol. 20, no. 65, 2018.

[3] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 3431–3440, 2015.

[4] F. Milletari, N. Navab, and S. A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *4th International Conference on 3D Vision*, 2016, pp. 565–571.

[5] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, 2017.

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci.*, vol. 9351, pp. 234–241, 2015.

[7] A. Kendall, V. Badrinarayanan, and R. Cipolla, "Bayesian SegNet: Model Uncertainty in Deep Convolutional Encoder-Decoder Architectures for Scene Understanding," in *The British Machine Vision Conference*, 2017.

[8] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation," *arXiv:1802.06955 [cs.CV]*, 2018.

[9] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 580–587, 2014.

[10] A. Shvets, A. Rakhlin, A. A. Kalinin, and V. Iglovikov, "Automatic Instrument Segmentation in Robot-Assisted Surgery Using Deep Learning," *17th IEEE Int. Conf. Mach. Learn. Appl.*, pp. 624–628, 2018.

[11] A. S. Fahmy, H. El-Rewaidy, M. Nezafat, S. Nakamori, and R. Nezafat, "Automated analysis of cardiovascular magnetic resonance myocardial native T1 mapping images using fully convolutional neural networks," *J. Cardiovasc. Magn. Reson.*, vol. 21, no. 1, pp. 1–12, 2019.

[12] A. Bratt *et al.*, "Machine learning derived segmentation of phase velocity encoded cardiovascular magnetic resonance for fully automated aortic flow quantification," *J. Cardiovasc. Magn. Reson.*, vol. 21, no. 1, pp. 1–11, 2019.

[13] H. Xue, E. Tseng, M. Fontana, J. Moon, and P. Kellman, "Inline Myocardial Perfusion flow mapping and Analysis: Powered by Gadgetron Inline AI," in *ISMRM 2019*, p. 2148.

[14] Z. Zhang, Q. Liu, and Y. Wang, "Road Extraction by Deep Residual U-Net," *IEEE Geosci. Remote Sens. Lett.*, pp. 1–5, 2018.

[15] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," *CVPR 2017*, pp. 5987–5995, 2017.

[16] M. S. Hansen and T. S. Sørensen, "Gadgetron: An open source framework for medical image reconstruction," *Magn. Reson. Med.*, vol. 69, no. 6, pp. 1768–1776, 2013.