INDIAN INSTITUTE OF TECHNOLOGY, KANPUR

SURGE-2023
Project Report

Revolutionizing Cardiac Image Segmentation with AI/ML Techniques

Submitted by
Abhishek Chaudhary
Application No.-2330609
Department of Chemical Engineering
IIT Kanpur, (U.P.) India

Under the guidance of

Prof. BV Rathish Kumar
Department of Mathematics and Statistics
IIT Kanpur, (U.P.) India

ABSTRACT

AI/ML has become the most widely used approach for cardiac image segmentation in recent years. In this study, cardiac image segmentation is performed using deep learning on common imaging modalities, including magnetic resonance imaging (MRI), computed tomography (CT), and ultrasound, and on major anatomical structures of interest (ventricles, atria, and vessels). Cardiac image segmentation is a crucial initial stage in various applications, as it involves dividing an image into meaningful regions that correspond to different anatomical structures. The key anatomical structures of interest in cardiac image segmentation typically encompass the LV, RV, left atrium (LA), right atrium (RA), and coronary arteries. Various deep learning methods like 2D FCN, 2D U-net + 3D U-net (ensemble), 2D U-net with cross entropy, etc. on cardiac MRI segmentation, especially ventricle segmentation, were applied to data sets, and results were analyzed. The existing method codes were also tested, and conclusions have been drawn. A large portion of these methods are designed for ventricle segmentation.

Keywords: Cardiac image segmentation, Machine learning (ML), MRI, CT, Ultrasound

OBJECTIVE

In general, these deep learning-based methods provide an efficient and effective way to segment particular organs or tissues (e.g., the LV, coronary vessels, scars) in different modalities, facilitating follow-up quantitative analysis of cardiovascular structure and function. Among these works, a large portion of these methods are designed for ventricle segmentation, especially in the MRI and ultrasound domains. The objective of ventricle segmentation is to delineate the endocardium and epicardium of the LV and/or RV. These segmentation maps are important for deriving clinical indices, such as left ventricular end-diastolic volume (LVEDV), left ventricular end-systolic volume (LVESV), right ventricular end diastolic volume (RVEDV), right ventricular end-systolic volume (RVESV), and EF.

FUNDAMENTALS OF DEEP LEARNING

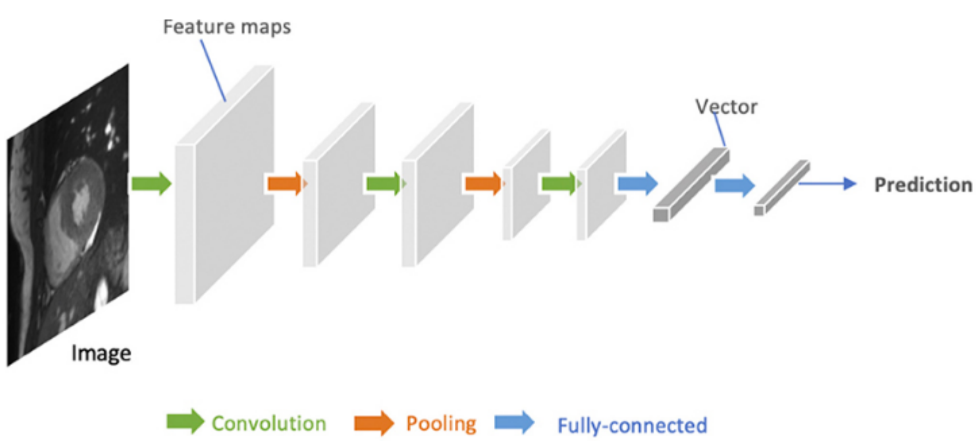
Deep learning models are deep artificial neural networks. Each neural network consists of an input layer, an output layer, and multiple hidden layers.

1. **Convolutional Neural Networks (CNNs)**:

In this part, we will introduce convolutional neural network (CNN), which is the most common type of deep neural networks for image analysis. CNN have been successfully applied to advance the state-of-the-art on many image classification, object detection and segmentation tasks.

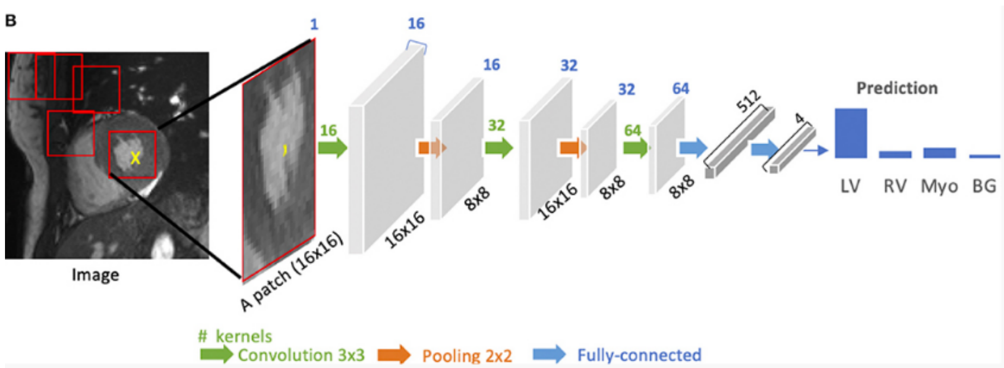
Generic architecture of Convolutional Neural Networks (CNN)

CNN takes a cardiac MR image as input. It learns hierarchical features through a stack of convolutions and pooling operations. Spatial feature maps are generated from these operations. The spatial feature maps are then flattened and reduced into a vector using fully connected layers. The vector output can take different forms depending on the task: Probabilities for a set of classes (image classification). Coordinates of a bounding box (object localization). Predicted label for the center pixel of the input (patch-based segmentation). Real value for regression tasks (e.g., left ventricular volume estimation).



Patch-based segmentation method based on a CNN classifier:

The CNN takes a patch (subset) of the image as input. The CNN outputs probabilities for different classes, with the class having the highest score representing the prediction for the center pixel of the patch. By forwarding patches located at different locations into the CNN for classification, a pixel-wise segmentation map for the entire image can be obtained. The classes in this specific example are LV (left ventricle cavity), RV (right ventricle cavity), BG (background), and Myo (left ventricular myocardium). Each convolution kernel used in the CNN is a 3x3 kernel with a stride of 1 and padding of 1, producing an output feature map with the same dimensions as the input. The number of channels of the feature maps is indicated by the blue number at the top.

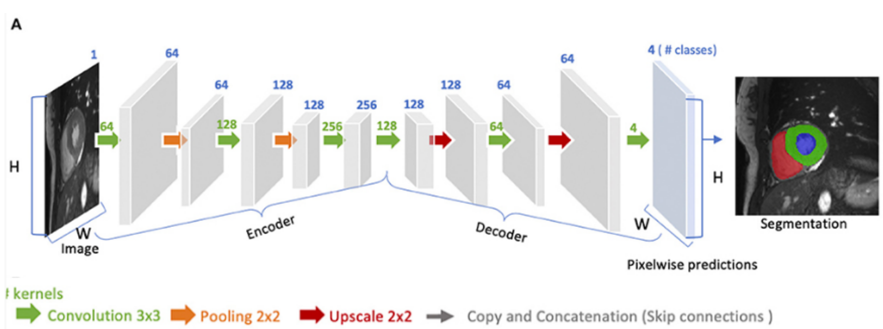


2. **Fully Convolutional Neural Networks (FCNs):**

FCNs are designed to have an encoder-decoder structure such that they can take input of arbitrary size and produce the output with the same size.

Architecture of a Fully Convolutional Neural Network (FCN):

The FCN takes the entire image as input. It learns image features through an encoder. The encoder gradually recovers the spatial dimension through a series of upscaling layers in the decoder. The FCN produces pixel-wise probabilistic maps, divided into four classes: left ventricle cavity (blue region), left ventricular myocardium (green region), right ventricle cavity (red region), and background. The final segmentation map is obtained by assigning each pixel to the class with the highest probability. An application of this FCN-based cardiac segmentation can be found in Tran's work.

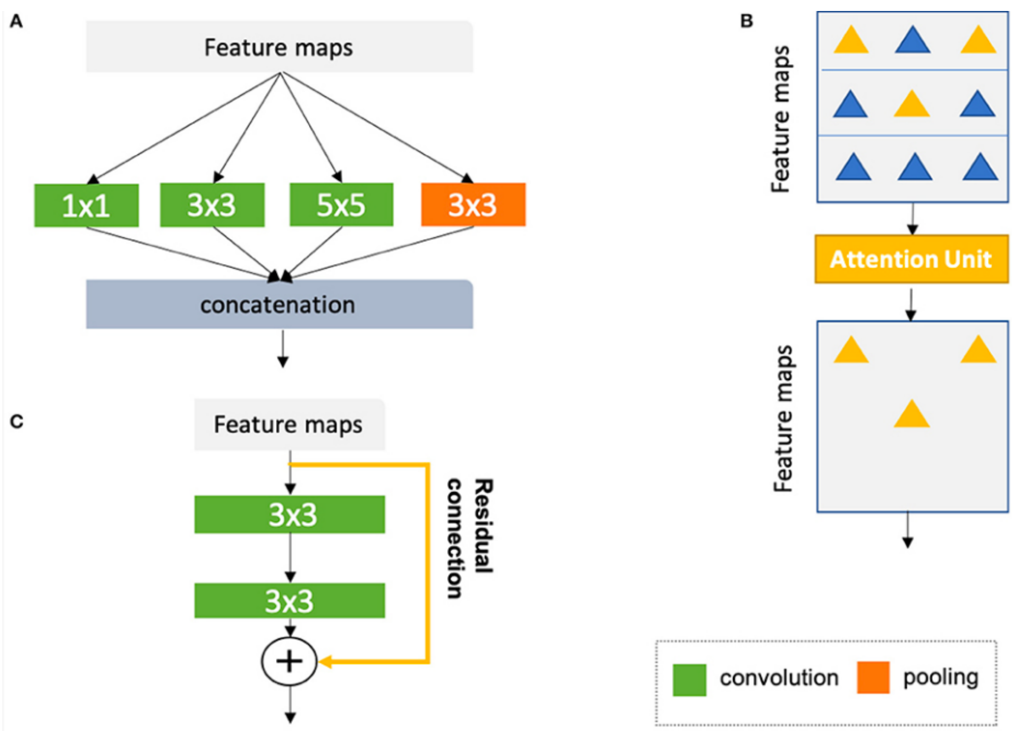


3. **Advanced Building Blocks for Improved Segmentation:**

These techniques are:

- Advanced convolutional modules for multi-scale feature aggregation
- Adaptive convolutional kernels designed to focus on important features
- Interlayer connections designed to reuse features from previous layers

<u>Generic Architecture of an autoencoder:</u>



A. Naive version of the inception module In this module, convolutional kernels with varying sizes are applied to the same input for multi-scale feature fusion. On the basis of the naive structure, a family of advanced inception modules with more complex structures have been developed.

B. Schematic diagram of the attention module The attention module teaches the network to pay attention to important features (e.g., features relevant to anatomy) and ignore redundant features.

C. Schematic diagram of a residual unit The yellow arrow represents a residual connection which is applied to reusing the features from a previous layer. The numbers in the green and orange blocks denote the sizes of corresponding convolutional or pooling kernels

4. **Training Neural Networks:**

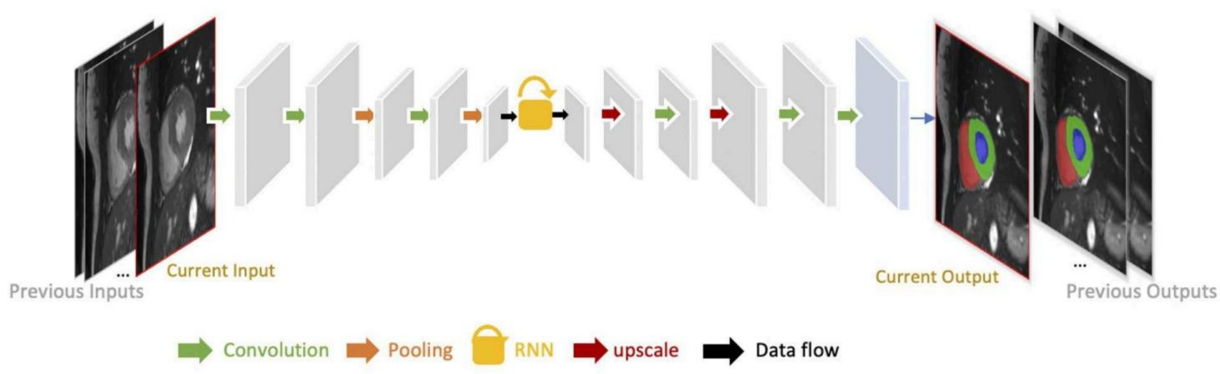
- Standard training process requires a dataset that contains paired images and labels {x, y} for training and testing, an optimizer (e.g., stochastic gradient descent, Adam) and a loss function to update the model parameters.
- Common Loss Functions
- Reducing Overfitting
  - Weight regularization,Dropout, Ensemble learning,Data augmentation,Transfer learning,Evaluation Metrics
  - Evaluation Metrics
  - volume-based metrics (e.g., Dice metric, Jaccard similarity index)
  - surface distance-based metrics (e.g., mean contour distance, Hausdorff distance)
  - clinical performance metrics (e.g., ventricular volume and mass).

THEORY AND METHODOLOGY

<u>Cardiac MR Image Segmentation:</u> Cardiac MRI is a non-invasive imaging technique that can visualize the structures within and around the heart. Ventricle Segmentation Atrial Segmentation Scar Segmentation Aorta Segmentation Whole Heart Segmentation

In this section, we will describe and discuss some methods for different applications in detail.

Vanilla FCN-based segmentation: The FCN architecture consists of an encoder network, which processes the input image and extracts hierarchical features, and a decoder network, which upsamples the features to generate dense pixel-wise predictions. The FCN produces pixel-wise probabilistic maps, divided into four classes: left ventricle cavity (blue region), left ventricular myocardium (green region), right ventricle cavity (red region), and background. The final segmentation map is obtained by assigning each pixel to the class with the highest probability.



Introducing spatial or temporal context: One drawback of using 2D networks for cardiac segmentation is that these networks work slice by slice, and thus they do not leverage any inter-slice dependencies. As a result, 2D networks can fail to locate and segment the heart on challenging slices, such as apical and basal slices where the contours of the ventricles are not well-defined. To address this problem, a number of works have attempted to introduce additional contextual information to guide 2D FCN. This contextual information can include shape priors learned from labels or multi-view images

Applying anatomical constraints: Another problem that may limit the segmentation performance of both 2D and 3D FCNs is that they are typically trained with pixel-wise loss functions only (e.g., cross-entropy or soft-Dice losses). These pixel-wise loss functions may not be sufficient to learn features that represent the underlying anatomical structures. Therefore, several approaches focus on designing and applying anatomical constraints to train the network to improve its prediction accuracy and robustness.

Multi-task learning: Multi-task learning has also been explored to regularize FCN-based cardiac ventricle segmentation during training by performing auxiliary tasks that are relevant to the main segmentation task, such as motion estimation, estimation of cardiac function, ventricle size classification, and image reconstruction. Training a network for multiple tasks simultaneously encourages the network to extract features that are useful across these tasks, resulting in improved learning efficiency and prediction accuracy.

Multi-stage networks: By applying neural networks in a multi-stage pipeline that breaks down the segmentation problem into subtasks. For example, a region-of-interest (ROI) localization network followed by a segmentation network. Likewise, a network called OmegaNet consists of a U-net for cardiac chamber localization, a learnable transformation module to normalize image orientation, and a series of U-nets for fine-grained segmentation. By explicitly localizing the ROI and rotating the input image into a canonical orientation, the proposed method better generalizes to images with varying sizes and orientations.

Hybrid segmentation methods: Another stream of work aims at combining neural networks with classical segmentation approaches, e.g., level-sets, deformable models, atlas-based methods, and graph-cut based methods. Here, neural networks are applied in the feature extraction and model initialization stages, reducing the dependency on manual interactions and improving the segmentation accuracy of the conventional segmentation methods deployed afterwards.

While these hybrid methods demonstrated better segmentation accuracy than previous non-deep learning methods, most of them still require iterative optimization for shape refinement. Furthermore, these methods are often designed for a particular anatomical structure.

RESULTS

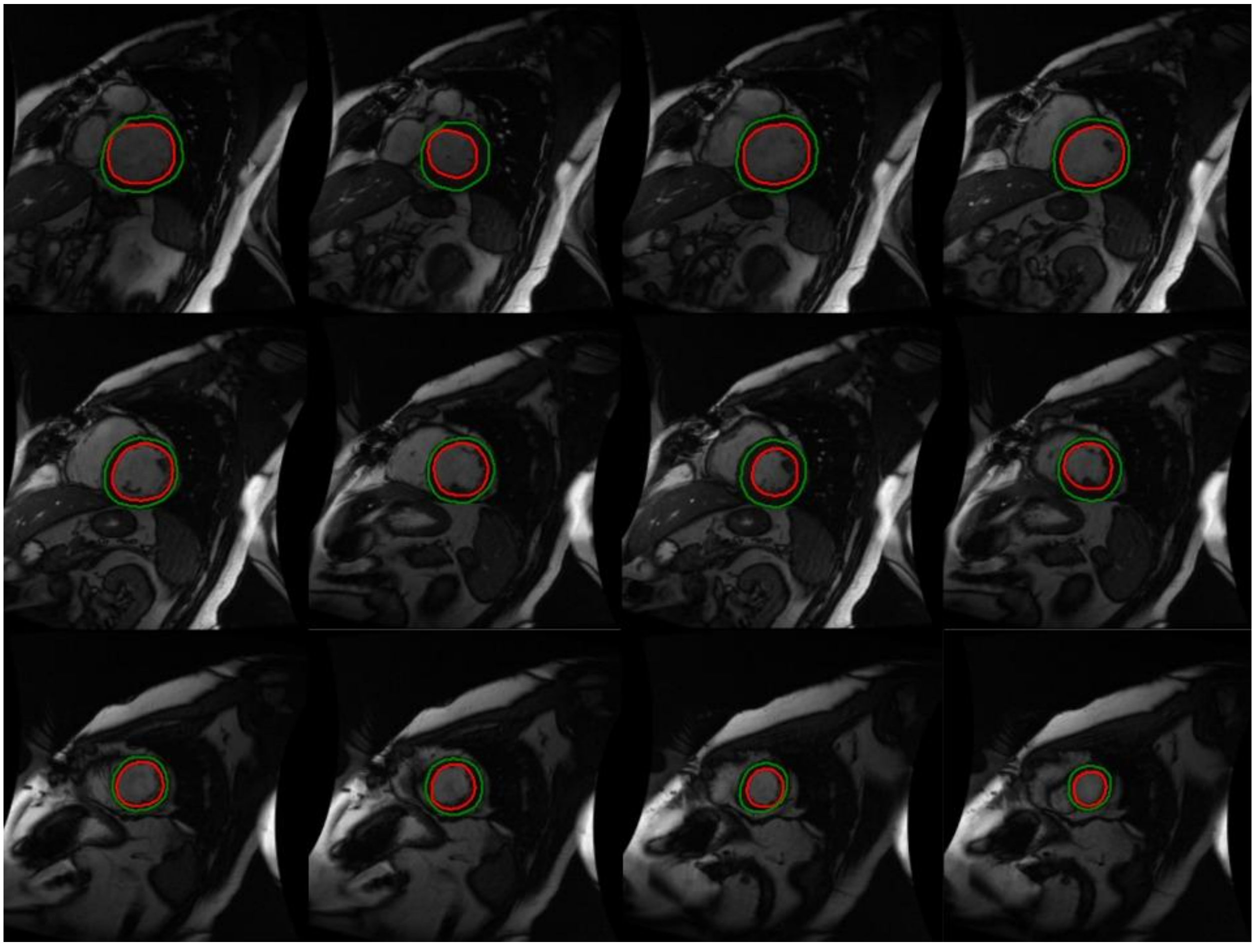


Fig 1: FCN segmentation result of an example test case in the Sunnybrook dataset for both ED and ES phases. Colors: red – endocardium; green – epicardium.

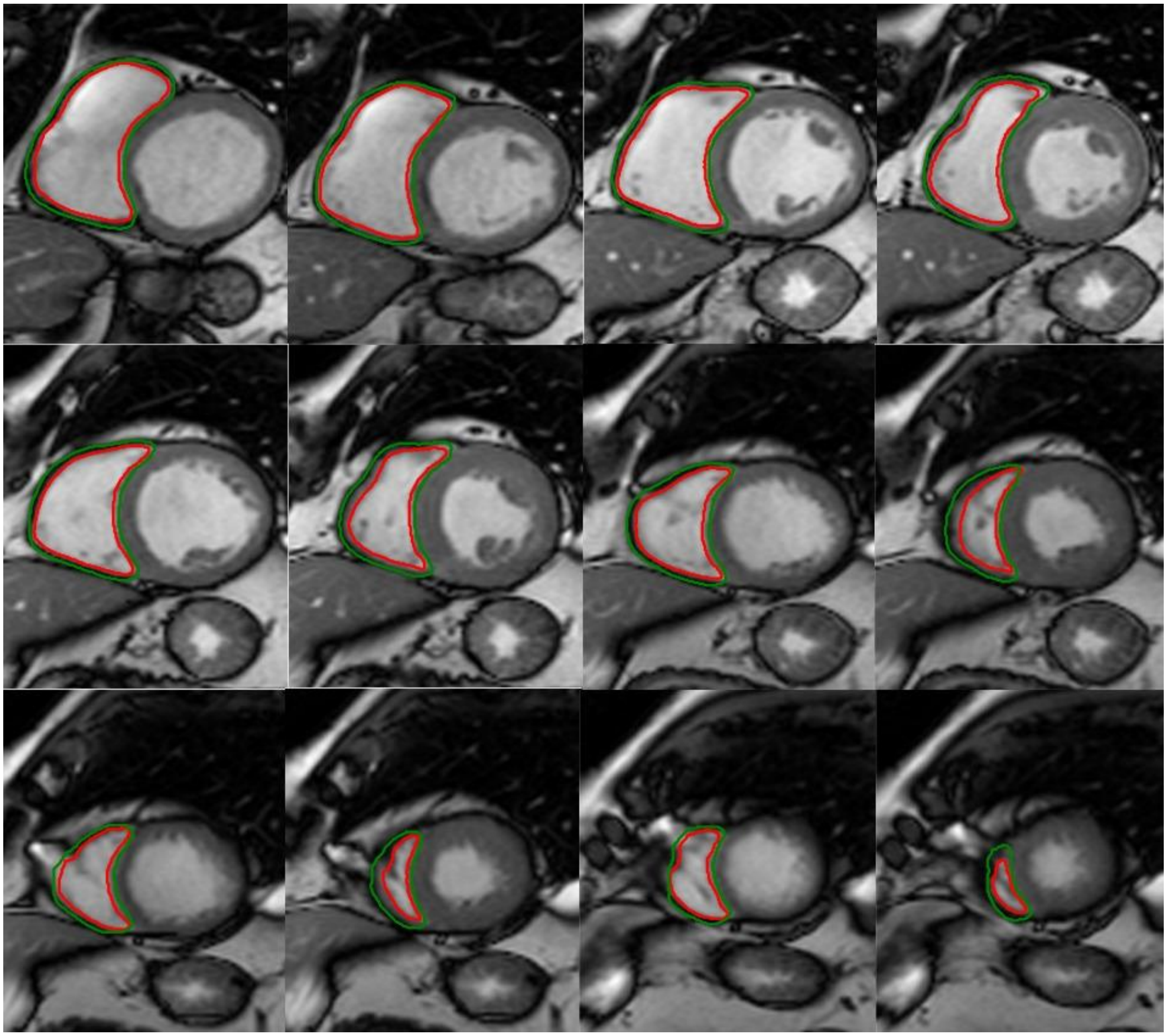


Fig 2: FCN segmentation result of an example test case in the RVSC dataset for both ED and ES phases. Colors: red – endocardium; green – epicardium.}

DISCUSSION AND CONCLUSION

In this paper, we demonstrated the utility and efficacy of a fully convolutional neural network architecture for semantic segmentation in cardiac MRI. We showed that a single FCN model can be trained end-to-end to learn intricate features useful for segmenting both the left and right ventricle. Comprehensive empirical evaluations revealed that our FCN model achieves state-of-the-art segmentation accuracy on multiple metrics and benchmark MRI datasets exhibiting real-world variability in image quality and cardiac anatomical and functional characteristics across sites, institutions, scanners, populations, and heart conditions. Moreover, the FCN model is fast, and can run on commodity compute resources such as the GPU to enable cardiac segmentation at massive scales.

An interesting conclusion is that the target image type can affect the choice of network structures (i.e., 2D networks, 3D networks). For 3D imaging acquisitions, such as LGE-MRI and CT images, 3D networks are preferred, whereas 2D networks are more popular approaches for segmenting cardiac cine short-axis or long-axis image stacks. One reason for using 2D networks for the segmentation of short-axis or long-axis images is their typically large slice thickness (usually around 7–8 mm), which can be further exacerbated by inter-slice gaps. It is well-known that training 3D networks is more difficult than training 2D networks. In general, 3D networks have significantly more parameters than 2D networks. Therefore, 3D networks are more difficult and computationally expensive to optimize, as well as prone to overfitting, especially if the training data is limited. As a result, several researchers have tried to carefully design the structure of the network to reduce the number of parameters for a particular application and have also applied advanced techniques (e.g., deep supervision) to alleviate the overfitting problem. For this reason, 2D-based networks (e.g.,2D U-net) are still the most popular segmentation approaches for all three modalities

REFERENCES

- [1604.00494] A Fully Convolutional Neural Network for Cardiac Segmentation in Short-Axis MRI - https://arxiv.org/abs/1604.00494
- [1709.04496] An Exploration of 2D and 3D Deep Learning Techniques for Cardiac MR Image Segmentation - https://arxiv.org/abs/1709.04496
- [1707.00587] Automatic Cardiac Disease Assessment on cine-MRI via Time-Series Segmentation and Domain Specific Features - https://arxiv.org/abs/1707.00587
- [1804.09400] 3D Consistent & Robust Segmentation of Cardiac Images by Deep Learning with Spatial Propagation - https://arxiv.org/abs/1804.09400

ACKNOWLEDGEMENT

I would like to extend my heartfelt thanks to my project mentor, Prof. BV Rathish Kumar, for his invaluable guidance and support throughout this journey. Their expertise, dedication, and encouragement have been instrumental in shaping my project and enhancing my skills. I am truly grateful for their patience, insightful feedback, and willingness to share their knowledge.

Additionally, I would like to express my sincere gratitude to the Surge program for providing me with this incredible opportunity. The program has not only enabled me to work on a challenging project but has also exposed me to a vibrant community of like-minded individuals. I am thankful for the resources, networking opportunities, and learning experiences that the Surge program has offered me, allowing me to grow both personally and professionally.