Machine Learning Nanodegree Capstone Project Proposal

Sathya Prakash

27 May 2017

Domain Background

Cardiac MRI is a diagnostic imaging tool to Image heart. With increase in Cardio Vascular diseases there is a need to know the condition of the heart regularly. Ejection fraction is a measurement of the percentage of blood leaving heart each time it contracts. During each heartbeat pumping cycle, the heart contracts and relaxes. When heart contracts, it ejects blood from the two pumping chambers (ventricles). When heart relaxes, the ventricles refill with blood. No matter how forceful the contraction, it can never pump all of the blood out of a ventricle. The term "ejection fraction" refers to the percentage of blood that's pumped out of a filled ventricle with each heartbeat [1].

Ejection Fraction is an important diagnostic metric to assess the condition of heart. Arriving at ejection fraction from a set of CINE MRI Images currently is a manual process. It requires expertise of clinicians to compute the ejection fraction manually.

Rudra et.al [4] from Kings' college London have applied Recurrent Fully Convolutional Networks to classify the Left ventricle from other areas surrounding it.

Few Projects have also been done in cs231n class of Stanford university. Taman [5] from cs231n has achieved a good performance using convolutional Neural Networks.

The Accuracy achieved by automatic Ejection Fraction Calculation are not still in the range which is acceptable for a clinical use. This remains active area of research. Arterys [6] a GE Healthcare company is also looking at commercializing the same.

Problem Statement

To estimate the Ejection fraction automatically from a given set of CINE MRI Images.
Convolutional Neural Networks have become very popular in performing Image Classification. The goal of this project is to train a Convolutional Neural Network to segment the regions of Left Ventricle and utilize this to determine the ejection fraction.

Dataset and Inputs

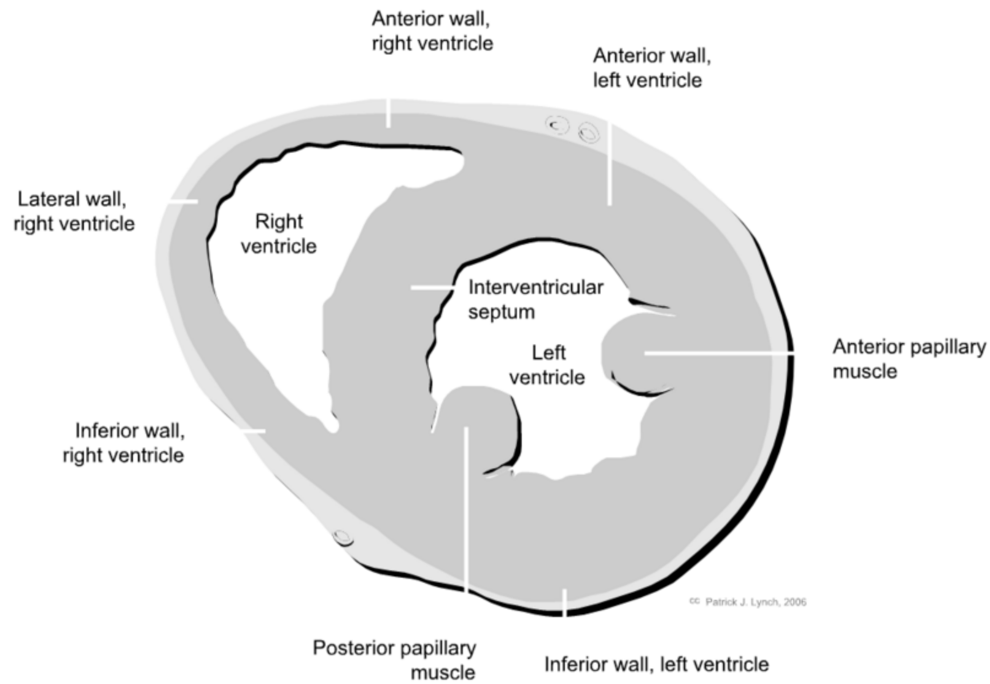
This project uses publicly available data from the following sources

1) **Data Science Bowl 2015 – Cardiac MRI Data set [2]**
2) **Sunny Brook Cardiac Data Set [3]**

Data consists of hundreds of cardiac MRI Images in Dicom Format. 2D CINE MRI Images contain 30 Images for a heart beat. As part of Training data systolic and diastolic volumes are provided. (Ejection Fraction can be easily computed using these).

Sunny Brook Cardiac train dataset also has contours representing the Left Ventricle.

Below is an image of how heart will look in a short axis view



Sunnybrook dataset consists of 45 sets of CINE MRI Images across various pathologies. The individual exams have been pre-grouped into training, validation and online testing subsets. Each subset contains 15 cases of which 4 heart failure with infarction (HF-I), 4 heart failure without infarction (HF), 4 LV hypertrophy (HYP) and 3 healthy subjects.

In all the images, the slice thickness is 8mm, the inter-slice gap is 8mm, the field of view is 320mm × 320mm and the pixel size is 256 ×256. In all 45 samples, LV endocardial contours were drawn by an experienced cardiologist by taking 2D slices at both the end-systolic and end-diastolic phases, and then independently confirmed by a second reader. The manual segmentations will be used as ground truth for the evaluation of the proposed models.

In Data Science Bowl 2015 , training set has 30 2D Images acquired in a single heart beat. Images are provided in DICOM Format , i.e it contains anonymized patient information like age heart rate etc. There is no contour provided as ground truth for LV segmentation. Data of 500 Patients provided. Images are of a maximum of 256 x 256 size.

Solution Statement

Solution employs convolutional neural network to train to segment the left ventricle from provided training dataset.

The Core model will employ multiple layers of CONV-RELU-MAXPOOL layers to build the network. In training model various set of cost functions including but not limited to cross-entropy loss will be used.

To implement Tensor Flow Library will be utilized.

Benchmark Model

Contours of left ventricle are available for test set of sunnyBrook Dataset also ejection fraction is provided for the test set. That will act as a benchmark model.

As part of this project I will aim to reach top 10% of Leader board in the Data Science Bowl 2015 competition.

Evaluation Metrics

I plan to evaluate the model on two tasks.

First, run it on the test set observations with ground-truth contours from the Sunnybrook data and compare the pixel-level predictions. This is the most direct test of the model. This tests how well we are able to segment/identify left ventricle.

Second, apply it to the Data Science Bowl 2015 data to compute left ventricle volumes for patients and compute its accuracy.

Project Design

Problem of finding ejecting fraction is decomposed into two problems

Identifying Left Ventricle in MRI Images

This is planned to be accomplished by using labelled data set provided in [3].A network featuring many layers of small convolutional filters along with ReLu nonlinearities and Pooling layers are planned to be used throughout.

In the final layer a Logistic function is planned to be applied to find the probability of each pixel being Left Ventricle or not. An appropriate Loss function has to be determined to train the network. Also Regularization will also be implemented in order to avoid overfitting.

Different Preprocessing needs to be performed on the MRI Images before providing them as inputs. The Exact Pre Processing will be determined during the data exploration stage.

As part of training the network a mechanism needs to be arrived at for determining hyper parameters that give better segmentation of the left ventricle. Once the Pixels belonging to left ventricle are identified then the next task would be to compute the volume of the blood flow for systolic phase and diastole phase to determine the ejection fraction.

Once the model is fine-tuned using the validation test will move on to the step of testing based on the completely new data that the model has not seen so far.

By getting access to the results of the Data Science Bowl 2015 [3] I plan to estimate the efficacy of the proposed design. At this point of time I'm not very sure if I can get access to the results of the test dataset in [3].

References

[1] http://www.mayoclinic.org/ejection-fraction/expert-answers/faq-20058286
[2] https://www.kaggle.com/c/second-annual-data-science-bowl/data
[3] http://www.cardiacatlas.org/studies/sunnybrook-cardiac-data/
[4] https://arxiv.org/pdf/1608.03974.pdf
[5] http://cs231n.stanford.edu/reports/2016/pdfs/317_Report.pdf
[6] https://arterys.com/about