

Automatic Left Ventricle Segmentation in Ultrasound

Ezequiel Ortiz

2018

1 Abstract

2 Acknowledgements

3 Intro

3.1 Motivation

3.2 Metrics

Assessing the function of the heart is crucial for detecting pathology as well as determining a diagnosis and prognosis for a patient. There are many heart metrics that physicians can measure, but two of the most common are left ventricular ejection fraction(LVEF) and left ventricular mass(LVM). Before the advent of these metrics, heart rate and blood pressure were the mainstays of cardiac function assessment. Without understating the importance of BP and HR, LVEF and LVM are the gold standard for cardiac We are focused on the left ventricle specifically because it is responsible for pumping blood throughout the body.

3.3 Ejection Fraction

LVEF is a ratio of end diastolic volume(EDV) to end systolic volume(ESV) to indicate how much blood the heart is pumping per beat.

EF it is calculated according to eq 2. Eq. 1

$$SV = EDV - ESV$$

Eq. 2

$$EF = \frac{SV}{EDV} * 100$$

Pinning down what is normal LVEF is not trivial. Everyone is different and LVEF that is healthy for one patient might be unhealthy for another. LVEF can also change depending on the physical state of the patient when the measurement is made. M. E. PFISTERER et al set out to determine ranges of normal LVEF by assessing the mean LVEF within a population of 1200 patients. The lower limit of normal LVEF values for men has been estimated to be $51.1 \pm 4.2\%$ with the upper limit of normal being $76.6 \pm 3.8\%$. These values were assessed with echocardiogram while the patient was supine and at rest. LVEF was also assesed with the more accurate but invasive x-ray angiography, and no signifigant differences were found. There were no signifigant differences in LVEF between sexes or age groups, except for patients over 60 who had slightly elevated LVEF levels. Patients that were able to exercise then reached 85% of their maximal HR before their LVEF was measured. On average, Men had an LVEF increase of 10.5% while women only saw an increase of 5.3%. Young patient LVEF increased about 6% more than older patients [1]. These findings

show that there is no bright line to determine what is normal or abnormal in LVEF.

Many variables affect LVEF, this obfuscates its relation to cardiac health. More context of the condition of the rest of the heart is required in order to make LVEF more useful.

Athletes often have LVEF values that below normal even though their hearts are healthy and functional [2]. And LVEF can be within normal values while the function of the heart is impaired. This is because EF only depends EDV and ESV and does not take into account the time duration or the geometry of the contraction. Getting accurate measurements of left ventricle EDV and ESV is not trivial. The vast majority of LVEF measurements are carried out by imaging the heart [2]. [3] LVEF values can change depending on the imaging modality being used.

3.4 Left Ventricular Mass

LVM is a measure of how thick the myocardium of the left ventricle is. Hypertrophic cardiomyopathy(HCM), as indicated via elevated LVM(how elevated?) occurs when the left ventricle myocardium thickens to a point where heart function is impaired. There are elevated risks of dilated cardiomyopathy in athletes as the natural thickening of the ventricle walls due to exercise can mask the condition.

Dilated cardiomyopathy(DCM) as indicated by lower than normal LVM(again what is normal) occurs when the left ventricle myocardium thins and stretches which compromises the strength and function of the left ventricle.

In conjunction, LVEF and LVM can help inform physicians on the efficiency of the heart and the reasons behind that level of function. There are many methods both invasive and noninvasive that can be used to measure LVEF. Noninvasive measurement techniques are always preferred if they offer similar performance to invasive methods. Imaging the heart has the ability of providing detailed structural information of the heart.

3.5 Cardiac Imaging

The most common imaging modalities used image the heart are Magnetic resonance imaging(MRI), x-ray computed tomography(CT), and ultrasound(US).

MRI employs strong magnetic fields and particle spin to MRI offers unparalleled contrast between different biological tissues. MRI is usually noninvasive but can require the use of contrast agents. Contrast agents are not usually used in cardiac MRI as the bloodpool and myocardium have good contrast, however gadolinium based agents can highlight structural differences within the myocardium. Once the heart has been scanned, it is relatively easy to calculate the interior volume of the left ventricle. For example, region growing methods are usually used to directly calculate the volume from the 3D scan. This approach is very accurate as no assumptions are made about the shape of the ventricle.

However, scanners are usually in short supply, require skilled operators, are expensive to run and maintain, and are slow to acquire images. When the subject of the image is moving, long image acquisition time can lead to motion artefacts. Patients must hold their breath while their heart is being imaged to avoid introducing such artefacts. Patients that are unable to hold their breath, or those whose heart moves while it pumps may be ill-suited for MRI.

CT imaging is not generally used for cardiac imaging as soft tissue does not provide sufficient contrast. The high energy x-rays used in CT are not heavily attenuated by low z elements found in soft tissue. Functional imaging of the heart with CT must utilize contrast agents [?]. High z compounds are injected into the bloodstream and subsequently imaged to give images of the blood-pool within the LV. CT imaging has excellent spatial and temporal resolution. Fast image acquisition means minimal motion artefacts when compared to MRI. However, the radiation dose and use of contrast agents make cardiac CT rare.

3.6 Ultrasound

Ultrasound is a modality unlike most others. High frequency sound waves are sent into the body, reflected, refracted and scattered, and picked up by the same transducer that produced the original sound pulse. These sound waves are reflected most strongly across boundaries of differing acoustic impedance. However, imaging the heart with ultrasound does present some problems. The ribcage encases the heart and a transducer that can fit in between the ribs must be used as bone is effectively opaque to the sound pulses. Once the transducer is in a position to image the left ventricle a view must be chosen. 2CH description. 4CH description.

3.7 Echocardiogram

3.8 Segmentation

3.9 Automatic Segmentation

3.10 Deep Learning

Neural nets. Neural networks are computational structures that mimic the function of biological neurons. Networks with sufficient complexity can model any mathematical function. The functional unit of neural networks is the neuron. Individual neurons, like the whole, will each receive an input and provide an output based on an internal activation function. Activation functions can vary in sophistication from simple step functions that provide a binary output based on a hard coded threshold, to sigmoid functions that organically map the input to a range of values between zero and one. The selection of activation functions will depend on the desired behavior of the network.

One of the simplest configurations of neurons into a network is the perceptron. We will examine a hand written digit classifier multi-layer perceptron (mpl)

to determine how neural networks make decisions. The structure of most networks consist of layers. In mpls, there is an input layer, hidden layers, and the output layer. The input layer in the case of the mnist digit image will simply be a vectorized image. The hidden layers contain the neurons that make the classification decision. The output layer will have a neuron for every class within the dataset so in our digit case we will have ten output neurons(0-9). The output layer will also be a softmax layer. The sum of all neuron values in a softmax layer must sum to one. For a given input each pixel value of our image will get sent to every neuron in the first of the hidden layers. These layers map the image input to a decision of what digit the network thinks it is seeing. If every neuron within the first hidden layer is getting input from every input layer neuron, then the output of every hidden layer neuron will be the same. In order to prevent this, we need to introduce the ideas of weights. Weights signify the strength of connections between neurons. If a neuron has been strongly associated with a three, in our digit classification example, then there will be neurons that fire strongly when the image provided is a three. The weights in any neural network determine the flow of information through the network and the final decision. In this classification example, each digit from zero to nine has different characteristics about it that make it easily recognisable to our brain as well as a trained neural network. A fully trained network is structurally identical to its untrained counterpart. It is the weights that dictate the strength of the connections and by extension the pathways responsible for the pseudo cognition.

The difference between the mlp and a deep network is simply the number of layers. The mlp by convention is limited to one hidden layer and any network with more than one hidden layer is considered deep. Deep networks have seen much more use than mpls recently as computer hardware advancements have allowed the training and implementation of networks with many more than two hidden layers. While the mlp simply took the pixel intensity levels as input and subsequently made decisions based purely on pixel intensity, more and more deep networks are utilising convolutions to extract relevant features from the input images. Convolution involves generating a new image from the input image. Each pixel of this new image is the result of an elementwise multiplication of an original image region and our mask and a . For example, a 3 by 3 averaging mask averages the intensities of a pixel and its eight neighbors to give the intensity of the output pixel. This operation is done for every spot that you can place the mask in the input image to generate the output. The nature of generating a single pixel output from a group of pixels means that there will be inherent

3.11 CNN

Padding is as simple as adding pixels to the edges of the input image to increase the resolution. Convolution naturally reduces the resolution of the output due to the thickness of the mask. There are many types of mask that can be used in convolutions and they produce a wide variety of output images that highlight different features of the original image. Sobel and laplacian filters make edges bright and flat surfaces dark. Averaging filters can mitigate noisy

images and produce blur. The size of mask can also be adjusted. We used an example of a 3x3 mask but larger masks such as 5x5 are also common. Masks are usually square with an odd width and height to ensure a middle pixel. All of the convolutions we have talked about are predetermined. A known mask is applied in order to produce a specific output. In deep convolutional neural networks, the masks are analagous to the mlp weights in that they will be defined through training. The generated masks do share some similarities with some of defined masks that we talked about earlier, because they require such functionality.

U-net. In order to segment the ventricle of the heart out of an ultrasound image, each and every pixel of the image must be classified as either within the ventricle or not.

4 Method

4.1 Dataset

The raw dataset was provided to me consists of pathological echocardiography images from 95 patients. Most patients had both a 2CH and 4CH view of their heart. The files are stored as zipped nifti images and are 2D grayscale videos of the heart through multiple diastole systole cycles. Constructing the final network that will segment cardiac ultrasound will be broken up into several successive steps.

1. Train a network to successfully segment ultrasound cone from the image
2. Train a network on 2CH bloodpool data
3. Train a network on 4CH bloodpool data
4. Train a network on both 2CH and 4CH bloodpool data
5. Depending on the success of combining 2 and 4CH data, train a network to segment the blood pool and epicardium

The purpose of segmenting the ultrasound cone from the rest of the image to make successive training easier. Eliminating image data that is irrelevant to the segmentation of the heart saves the network from having to learn that the extraneous information is so. The build up to the final network allows me to gradually build up the training dataset.

Once the dataset is complete, conversion into a form that can be used to train the U-nets is the same for 2CH and 4CH images. As this project was written within Google Colaboratory on account of the free cloud GPU, all of the image data was hosted on my personal Google Drive which was then mounted within the colab environment.

The images need to be converted into numpy arrays before they are used to train the network. The python library nilearn is used to convert the zipped nifti files into numpy arrays. The images are then resized to a resolution of 512 by 512 and the intensities are remapped from 0-255 to 0-1. The images are then collectively stored in a 4th order tensor with dimensions (number of images, X dimension, Y dimension, color channel). This allows for easy manipulation with Keras

The masks go through the same resizing conversion from nifty to numpy array. As the mask images can contain cone or ventricle masks, one layer of the masks must be chosen for training. This is accomplished via a thresholding operation to eliminate either the cone or ventricle mask. Once the correct mask configuration has been chosen, the masks can be resized and stored in a fourth order tensor as well.

dataset preprocessing The data was presented to me in the form of 96 folders, each corresponding to a certain patient. The names of these folders were all unique as they came from different hospitals at different times and were aquired by different machines and operators. Most patients had both two and four chamber images, with some having multiple of either. the naming convention for the individual images was as follows:

US_2CH.nii.gz
US_4CH.nii.gz

and if there were multiple of one view a two is appended to the end of the name as follows, US_2CH2.nii.gz

One of the first things that I did was to seperate the dataset into two and four chamber images. However, the image files would need to be renamed so that they could be told apart. The name of the directory containing the images was appended to the image filenames so that there would be no conflict of names and the image view sould still be identified by reading the name. An example of a filename in this form is as follows: KCL_GC-001-US_4CH2.nii.gz Once the images were somewhat sorted, manual segmentation could begin. Behind every competent neural net is a heap of painstakingly aquired data. To create a data point the blood pool and ultrasound cone from an individual frame of the nifti file must be segmented. To do this ITK-snap was used to both open the nifti file and produce the segmentations. In order to increase variance within the dataset and therfore robustness of the trained network, the chosen frames are at the general point of diastole and systole. Most all of the scans taken were of pathological hearts, so finding exact diastole and systole was challenging. Most easily distiguished features of the ecg within the echocardiogram were undistiguishale to our untrained eyes. We usually settled on two or threeframes from each scan that looked sufficiently different. If more data is needed at a later time, additional frames could be segmented. The product of the segmentation is

a mask that contains information on both the boundaries of the ultrasound cone and bloodpool of the left ventricle. As the ultrasound cone is quite simple to identify within the images, we only ended up producing 25 data points with the cone manually segmented. Whereas every single mask contained ventricle information. In order to train a network on the images, they needed to be a standard resolution. Initially this resolution was to be 512 by 512. 512 is a power of two in order to avoid non integer resolutions as the images are downsampled in the unet. The raw images in the dataset varied in resolution(TODO exact reses pls) with 512 by 512 being a rough mean. We later moved to a resolution of 800 by 800. This change was made based on two factors. We wanted to zero pad the images up to a resolution instead of cropping them or interpolating them to a smaller size. Interpolation would give a reproduction of our images at a different resolution while zero padding would encase the exact original image in zeros. While interpolating image data is common practice, the end goal of producing a measurement of LVEF in real world units places a higher degree of importance in preserving the pixel spacing and dimension information contained in the image headers. As the data generation script took shape, it was clear that keeping the original pixel spacing measurements would save lots of time(TODO make clearer) scaled from 0-1 instead of 0-255,standardized(TODO ask esther!)

training

validation

data generation

References

- [1] M. E. PFISTERER, A. BATTLER, and B. L. ZARET, “Range of normal values for left and right ventricular ejection fraction at rest and during exercise assessed by radionuclide angiocardiology,” *European Heart Journal*, vol. 6, pp. 647–655, 08 1985.
- [2] T. H. Marwick, “Ejection fraction pros and cons: Jacc state-of-the-art review,” *Journal of the American College of Cardiology*, vol. 72, no. 19, pp. 2360 – 2379, 2018.
- [3] N. H. M. Arrow, N. K. McAlister, and K. Butto, “Understanding cardiac “echo” reports,” *Canadian Family Physician*, vol. 52, pp. 869–874, 2006.