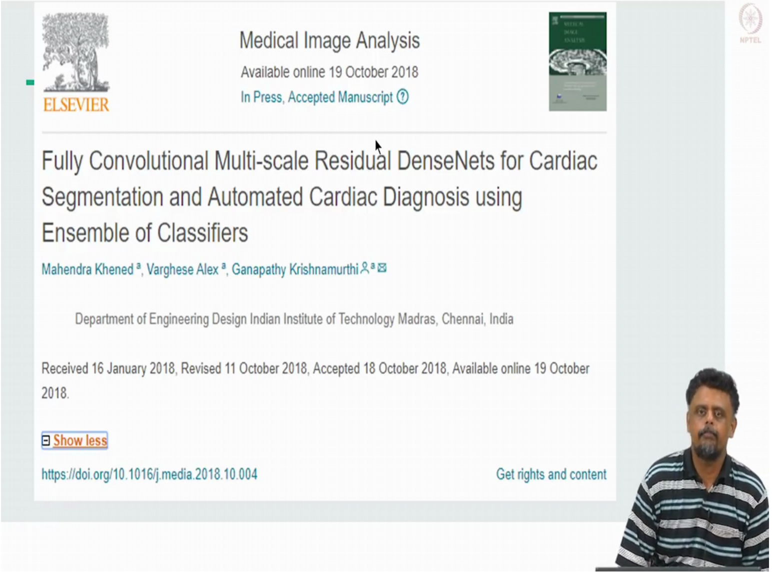
Machine Learning for Engineering and Science Applications

Professor Dr. Balaji Srinivasan
Department of Mechanical Engineering
Indian Institute of Technology, Madras

Applications: Cardiac MRI - Segmentation & Diagnosis

(Refer Slide Time: 0:20)



Medical Image Analysis
Available online 19 October 2018
In Press, Accepted Manuscript

Fully Convolutional Multi-scale Residual DenseNets for Cardiac Segmentation and Automated Cardiac Diagnosis using Ensemble of Classifiers

Mahendra Khened [a], Varghese Alex [a], Ganapathy Krishnamurthi [a]

Department of Engineering Design Indian Institute of Technology Madras, Chennai, India

Received 16 January 2018, Revised 11 October 2018, Accepted 18 October 2018, Available online 19 October 2018.

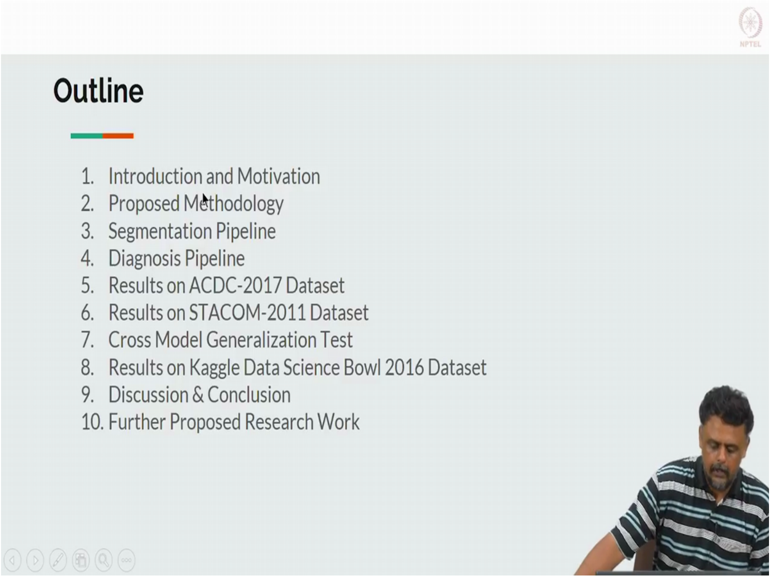
Show less

https://doi.org/10.1016/j.media.2018.10.004

Get rights and content

Hello and welcome back, so in this application week we will look at an application of deep neural network typically CNNs and a combination of machine learning algorithm classification of algorithm, a problem in medical imaging or medical image analysis. This is based on a paper out of lab which appeared in medical image analysis titled “Fully Convolutional Multi-Scale Residual Densenets for Cardiac Segmentation and Automated Cardiac Diagnosis Using Ensemble Of Classifiers” ok, so my Ph.D. students have co-authored this paper along with me ok, Verghese and Mahindra Khened. Mahindra will walk in trough now the code and some pieces of this algorithm later on, I will just give you an overview right now.

(Refer Slide Time: 1:02)

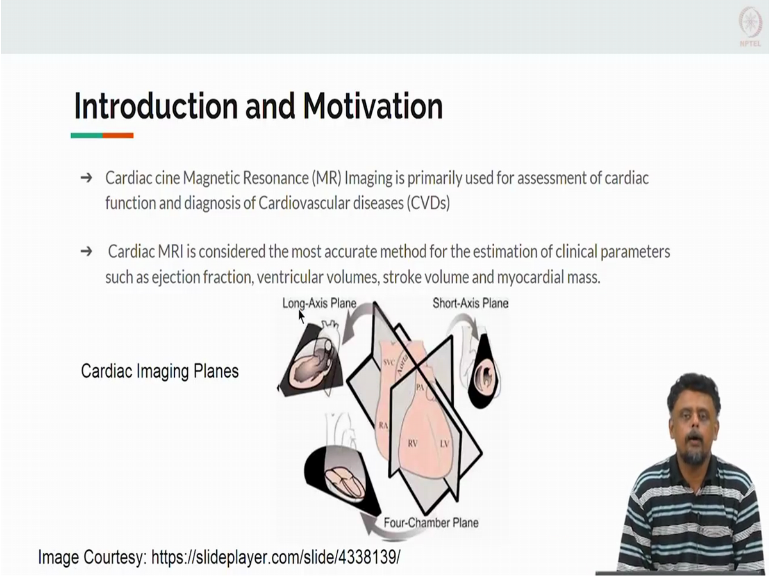


Outline

1. Introduction and Motivation
2. Proposed Methodology
3. Segmentation Pipeline
4. Diagnosis Pipeline
5. Results on ACDC-2017 Dataset
6. Results on STACOM-2011 Dataset
7. Cross Model Generalization Test
8. Results on Kaggle Data Science Bowl 2016 Dataset
9. Discussion & Conclusion
10. Further Proposed Research Work

This is the outline of what we are going to look at, so we will give you introduction and motivation to this particular problem because you might not be aware of what medical imaging technique are and what they are used for, et cetera. And then the proposed methodology, segmentation pipeline, and the rest about diagnosis of the pipeline and the code walk-through and some aspects of the data will be done by Mahindra okay.

(Refer Slide Time: 1:25)



Introduction and Motivation

- Cardiac cine Magnetic Resonance (MR) Imaging is primarily used for assessment of cardiac function and diagnosis of Cardiovascular diseases (CVDs)
- Cardiac MRI is considered the most accurate method for the estimation of clinical parameters such as ejection fraction, ventricular volumes, stroke volume and myocardial mass.

Cardiac Imaging Planes

Image Courtesy: https://slideplayer.com/slide/4338139/

So let us go to the introduction, so if you are all familiar with magnetic resonance imaging, if not then just look it up on the web or some resources, physics of how it operates, it is the entire course by itself, but we will just see for the sake of these slides you can guess what it is used for. Magnetic resonance imaging is one of the main tools for assessment of cardiac function okay, cardiac function means function of your heart ok. And it is also considered the most reliable method and people used at in the clinic when you go to the hospital and take an MRI of your heart, they can use that to figure out if there is something wrong with the function of your heart ok.

So let us see what MRI does, so what many of the modern image for imaging techniques do, which is CAT scan or in this case the MRI scan, it gives you sections of an (())(2:12) okay. So if you recall, if you go to hospital you lie down on a flatbed and then the bed moves into the gantry and you lie there for some time while the images are acquired. So it is like slices your it acquires slices of images that run through your cross-section of the body okay, so these are called referred to typically as axis slices.

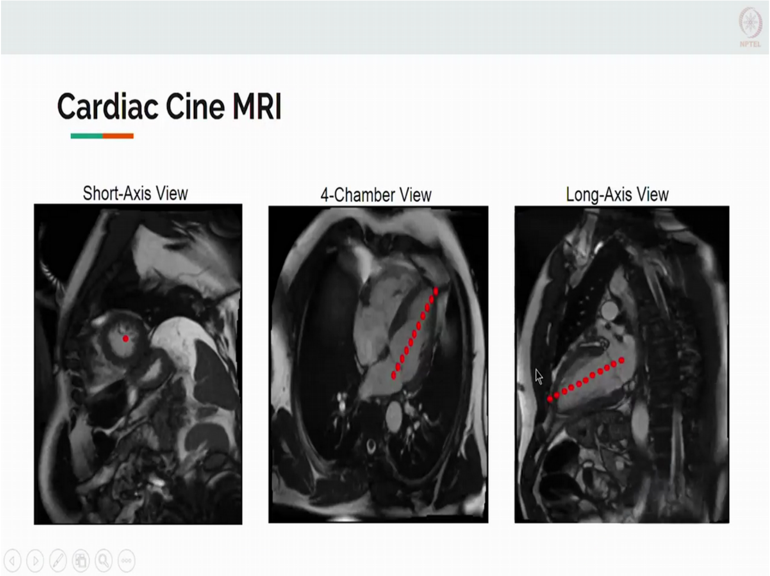
MRI is slightly more flexible, it can acquire images at random cuts so you can make a cut like this and look at the cross-section of it ok. So for the purpose of this particular paper we are going to look at the cardiac imaging plains. I also urge you to look up some 3-D models of the heart, lots of the atomic images are available online, but for this lecture we will just concentrate on this one. We typically acquire a volume of the heart, so the entire organ is covered by this can so slices through the heart are acquired and the way we slice the notation is important. So if we look here, what is going to (())(3:11) LV is the left ventricle. Left ventricle is the chamber in the heart that pumps out blood into your system ok, through the iota that is where the (())(3:23) blood comes out of your heart.

And if you cut and LV, it looks like kind of a overleaf if you think about it, and if you cut across its short axis those image sections are known as short axis images. And there are the long axis planes which cuts through the section of the heart, think of it like cutting this way okay, suppose a short axis which cuts this way right. And the other plane is a chamber plane which is basically the plane which will show you all compartments of the heart; the heart has 4 compartments or 4 chambers; right ventricle, left ventricle, right atrium and left atrium. So if you want a view that has all the 4 chambers, cross-section of all the 4 chambers in it that is called the 4 chamber plane ok.

So typically these are the 3 orientations at which the volume of the heart is used, so I can think of the image acquired as a volume, volume is a 3-D matrix ok. It is a grid 3-D grid and where it is pixelized of course because it is a reconstructed image, acquisition is discreet okay, the sampling is discreet so you get a discretised image 3-D discretised image and you can look at each plane of the 3-D volume as belonging to either a short axis, long axis or the

4 chamber plane ok, so we will be primarily working with the short axis plane, so we will see what that looks like ok.

(Refer Slide Time: 4:43)



Cardiac Cine MRI

Short-Axis View

4-Chamber View

Long-Axis View

So here is the picture an animation of the short axis view. So if you look at this, this is the compartment through which the blood actually goes through, the dark region here that is called the myocardium that is the one, the myocardium is the muscle ok the muscle which actually twists and compresses the heart, you can think of it that way and pumps the blood out of the heart into the system ok. So if you acquire the images of a (())(5:12) heart is not a stationery organ and since magnetic resonance imaging helps you acquire images very quickly, you can actually acquire heartbeat. Since you cannot sample different heart beats at (())(5:25) time intervals and put together this image.

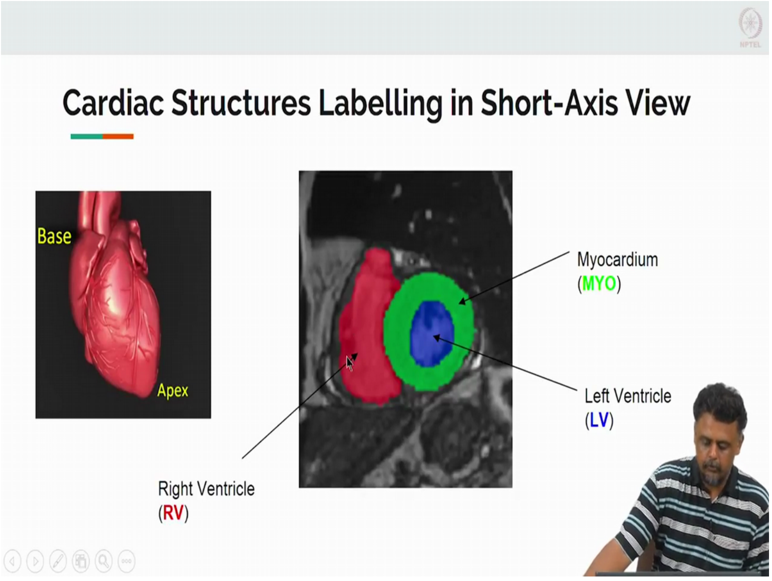
So the entire one heartbeat is one cardiac cycle like compression and then release that is acquired during that the entire stroke one stroke of the heart which is like it pumps the blood out that is called one cardiac cycle and the images are acquired throughout the cardiac cycle. So you have volumes, let us say for instance typical heartbeat is one second, you can end up acquiring 20 or 25 time points in that one second ok. In each second you will have the heart in a different state of compression or expansion in one cardiac cycle okay, you can think of it that way. So this plane on the left that you are looking at is the axial plane okay, you can see the cross-section this is a short axis cross-section ok.

The 4 chamber view is right there, again the left ventricle are long axis are shown right there and this shows all 4 compartments, you can see the heart, the section of the heart where all 4, the blood travels through all the 4 compartments before it gets pumped out of the system ok. And so that is the 4th chamber view right then and this is the long axis view when you can see around the length of the left ventricle if you can think of that way. Typically for all this cardiac MR imaging there are many diagnoses that are done by with cardiac MRI, but what is of most interest to physicians, clinically important factors are obtained by looking at the left ventricle because that is the one that pumps blood out into your system and any weakness there or any defects of pathology there is obtained by studying the left ventricle during the cardiac cycle ok.

So another terminology that you should be aware of is called the systole and diastole, so think of the diastole as when the left ventricle is build with the blood that is when the myocardium is let us say is relaxed a little bit and it fills the blood that is the diastole, the systole is when it is fully compressed that is when it is actually pumping it at out of the system, so these are the two, and the images are acquired between N-systole and N-diastole that is typically the terminology that is used. So just to reiterate it, let us say if your heartbeat is 60 beats per minute so you have one bit per second which corresponds to pumping out once, so one pumping action, so there are typically 20 to 30 image volumes the entire volume of the heart is acquired between that one cycle.

So if you run it like a movie that is what you get, if you put together if you take a particular cross-section of the heart and you take the same cross-section from every time point image and put it together, you can make a movie like this that is how this is done.

(Refer Slide Time: 8:15)

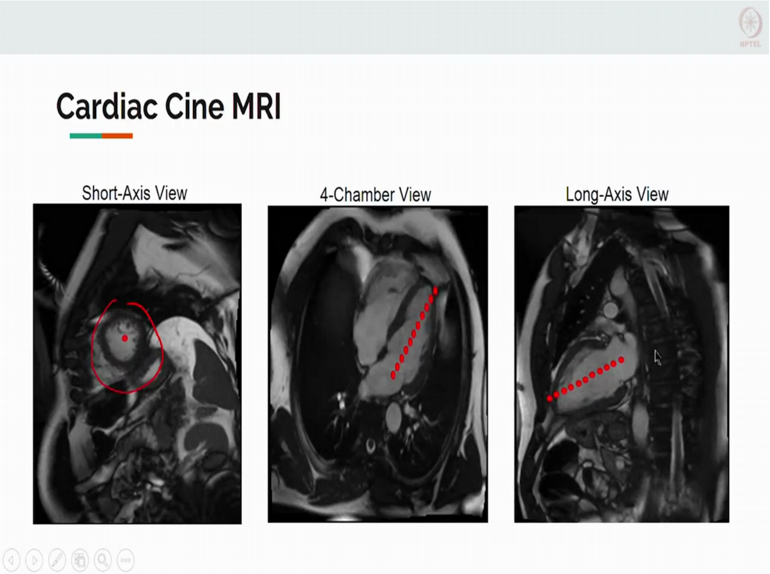


Cardiac Structures Labelling in Short-Axis View

There are some other terminologies that gives better view of the heart, so here you can see the base of the heart is basically, if you think about it is the top of the heart, it is standing up that is where the base is and the apex is that tip right there, so that is how is the direction if you think of it the base to apex or apex the basis how it is referred to as ok. And just concluding from the previous slide, if you look we have this region in green which is referred to as the myocardium, myocardium is the muscle that does the pumping okay. If there is something wrong with that muscle that is when you get all these heart problems like heart attack, cardiac arrest, things like that.

Left ventricle is the cavity which contains the blood okay, and right next to it is adjacent chamber is the right ventricle again highlighted in green.

(Refer Slide Time: 9:09)

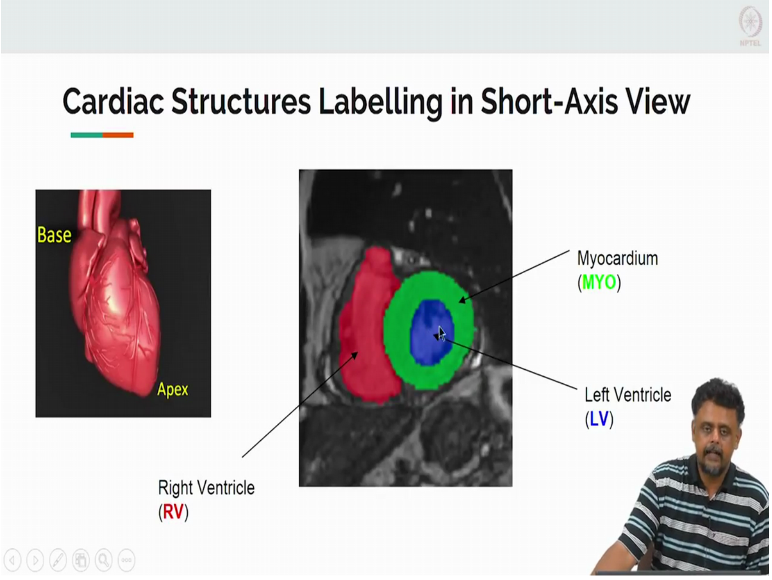


Cardiac Cine MRI

So if we go back to the previous slide, so if we look at this particular slice here, we can see the LV very clearly, this is just the left ventricle alone. And if you look at this slice right there, that is where the this is the left ventricle as I point out right next to it is the right ventricle that is the region, I just marked it just to show you of course. So I am not an expert in anatomy, I just learn some of it in the context of this problem okay, so you should if possible consult some radiologist or somebody who is a renowned medical student or somebody who has done medicine, he can probably point this way better than I can. In the conduct of the problem these are the anatomies that we will be concerned with.

So the dark region that you see use a different colour, maybe I will use green so this dark region here which I am going to share that is where the myocardium muscles, those are the muscles which pumps your blood out of the system okay.

(Refer Slide Time: 10:06)



Cardiac Structures Labelling in Short-Axis View

So that is being highlighted here, the left ventricle is the cavity that holds the blood the surrounding muscle that is the myocardium and adjacent chamber is called the right ventricle ok. So before we go any further so what is the challenge? Okay so what are we going to do, so a lot of times when the patient its image, doctors can the radiologist can just go through these images and can see the effect sometimes. A lot of time some quantitative analyses has to be done, so one of the things several factors that are calculated are the ejection fraction okay, you know if you want to look at the area of the myocardium or you know the size of the myocardium and also for all of those tasks where you need some quantitative markers to be obtained from the image, what you need is accurate segmentation.

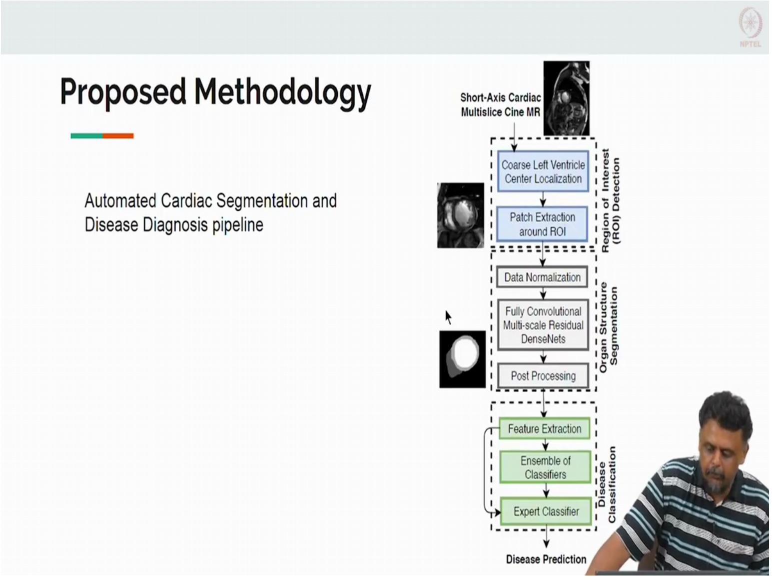
So this is what we use what we call semantic segmentation when we look at convolutional neural networks. So basically we have to label the pixels corresponding to the myocardium, label the pixels corresponding to the left ventricle and labelled the pixels corresponding to the right ventricle ok, so these things have to be accomplished automatically. So why is this so important to do it automatically you, why cannot just the radiologist go and do it? There are problems there will be many volumes in time so I will talk about that, but even if we consider volumes acquired at N-systole and N-diastole that is one friend N-diastole is we can think of when left ventricle is filled with blood and N-systole is when it is pumped out okay.

So if you look at it, so each volume will have hundreds of slices okay each volume will have hundreds of slices, and if you are also going to go through several time point then it means that you are going to have look at about thousand slices of information, thousand pictures if you want it that way, thousand pictures in which it has to go and manually mark its boundaries that is going to be prone to error and it is definitely going to be time consuming, I do not think he can just see more than a patient a day that is the case okay. So if in that case so in automatic techniques for figuring out this relationship for doing the segmentation right, so that is one thing. 2nd important thing is what do we do with all this quantitative measurements right? What do we do with it?

The idea is to predict disease okay, so a patient walks in may be is being referred to by some doctor by some cardiac specialist that may be used should have your heart scan by MRI, so he does the scan and the image comes out ok. Then maybe it is nice if you have a system that provides a diagnosis on its own like what kind of, there are several conditions we will see what those conditions are, in this case we are considering I think 4 conditions 4 to 5 conditions which can be actually inferred from the images themselves, can we do it automatically so that is the question. For instance, this is like saying if you look at the you know do a whole body scan of the patient and say you have cancer something like this, some diagnosis like that without having manual intervention so that is the idea.

So there are 2 folds; one is the automatic semantic segmentation as we know it, the 2nd is prediction of the disease in the heart based on the quantitative markers or values that we obtain from the segmentation ok, we will see what those values are later on in our talk.

(Refer Slide Time: 13:27)

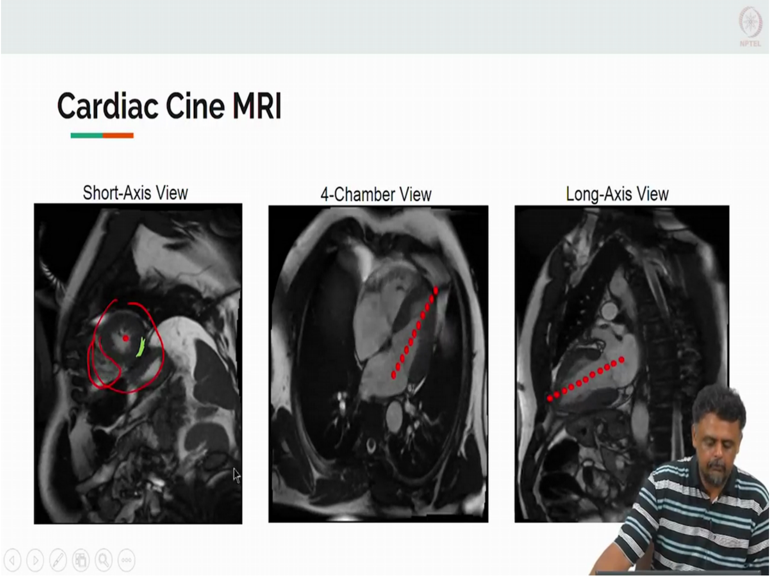


Proposed Methodology

Automated Cardiac Segmentation and Disease Diagnosis pipeline

So what is the pipeline for this work? Okay, so generally this as we have about 3 blocks okay, so the 1st one right here, the 1st one we are going to look at that is the preprocessing okay, it contains everything that some of the aspects of you know (())(13:45) also is contained in that so we will see what this preprocessing techniques exactly is, so the preprocessing helps us to actually zoom in on the region in the image that is of importance to us.

(Refer Slide Time: 14:05)

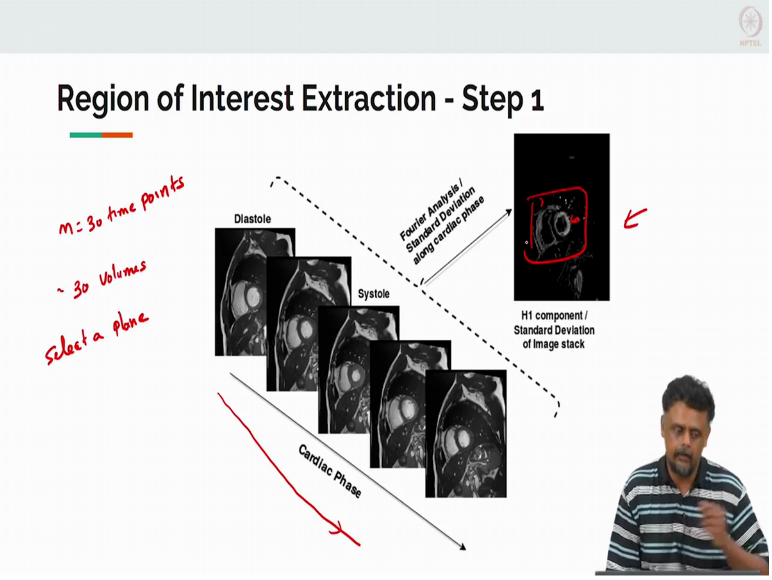


Cardiac Cine MRI

So why would that be important? So if we look at the picture here, see it is a pretty big image I think right and it is about 512 by 512, it can vary from 256 cross 256 to 512 by 512. Look at it only a small region here, this is the region of interest for us okay, this is where the heart is, cardiac muscles are and only we need to we need only this region. So to make our CNN's job easier so we do not want a very large input to our CNN, if he can just crop this region out in every image successfully and automatically that will be useful okay, so then that is what this particular piece does.

The 2nd piece again data (())(14:38) because it is part of the training process I guess of the neural network so we have a neural network portion here so that is our deep CNN we can call it okay and the output is the segmentation, we do some postprocessing of the segmentation, make sure that it is clean, the third-place here is a classification of disease ok, so these are the system. So we will go through each one of them, I will give you a brief overview and then we will see that somewhat the individual steps are in much more detail.

(Refer Slide Time: 15:18)



Region of Interest Extraction - Step 1

M = 30 time points

~ 30 volumes

Select a plane

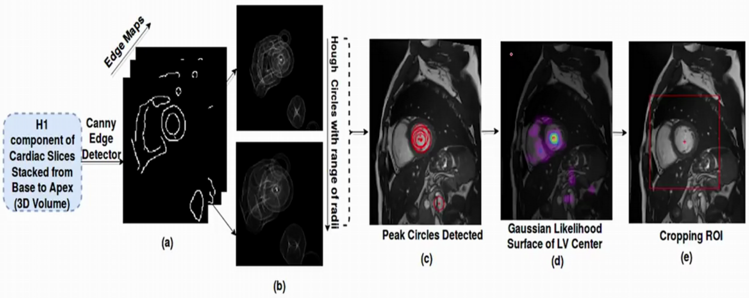
So the region of interest extraction, look at the step 1 so how is this accomplished right? So I have mentioned to you that if you have let say if you have 60 beats per minute that is your heart rate so you get about in a second you have 1 beat is completed and then we acquire about 23 or 20-30 volumes in that one beat right, that is the entire volume. So then that is useful because if we consider the same plane the same height, same cross-section in one of those time points let us say so for instant we acquire M = 30 time points, which corresponds to 30 volumes ok.

We select a plane, the scanner has some references so then we will be able to figure out in actually absolute terms where a particular Z value lies, where a particular coordinate lies, so we should be able to select that plane and if you arrange that plane in the time series that is what it is called, it is seen as a cardiac phase so different stages of compression or expansion of the heart is referred to as cardiac phase. So if you arrange them, you see a stack of images and we can take a standard deviation across that stack. So you can arrange them as stack and if you take one pixel you can look at that pixel through the stack and calculate the standard deviation.

So that standard deviation picture is what you see here, what is obvious here is that if you look at the edges right that has the high value of standard deviation not surprisingly because the heart is moving so it expect a much larger standard deviation near the edges where the myocardium is and that will help us actually localise the heart. So we can to standard deviation in threshold and the regions of large standard deviations would correspond to the moving parts of the image. So if not all algorithms are moving, there will be some slight distortion of the liver and all that because of the heartbeat but that is not too significant so the highest standard deviation would correspond to the moving anatomy which is the heart so that is the image that we get, the standard deviation image taken across different time points of the cardiac cycle or the cardiac phase.

(Refer Slide Time: 17:35)

Region of Interest Extraction - Step 2



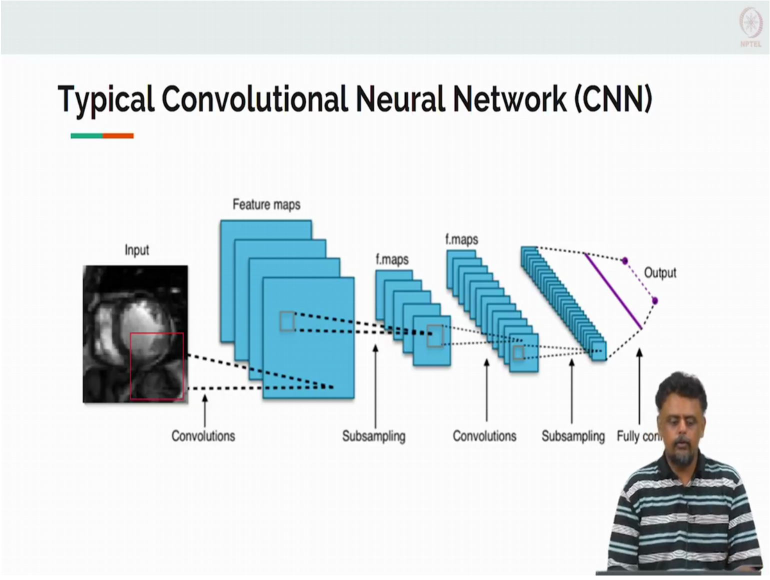
So once we have that we can do an edge detector on the standard division image okay, we can use the standard canny edge detector, if you do not know what a canny edge detector is, you can search about it in the subject, Matlab has canny edge detector inbuilt so you can just look for it in the help section and try it on some standard images that are provided. So once we do this edge detection, we can we can then use, now you see that the circular objects are quite well-defined when you do the (())(18:09) or edges are clean. On top of that you do what is called hough transform ok, so the hough transform again we have looked this up, we are not going to go through all of these techniques as it is not image processing course.

So hough transform helps detect circles, so this is circular hough transform that we are using has detect circles and images, typically if you have black and white image like the one we have generated here, you can use it to automatically detect circles. Now if you are wondering why that is so important, just think about how you would actually detect a circle in an image, it is actually a nontrivial problem to do okay. So we will use a hough transform today itself, and once we have the hough transform to detect the circle then we can localise the left ventricle centre of the left ventricle through some statistical techniques and then have a box around it that is the prop, so we have fixed a standard box size and we take that out okay.

So do all the images we will use typically for training and testing and validation. On the other hand, we can say that well the CNN is supposed to figure this all (())(19:15) that is also possible, if you do not want to do this crop preprocessing which involves localising the left ventricle, you can actually feed the direct image all the entire image directly to be seen as input. I am just going to take it when we will study more typical problems to solve because there is so much (())(19:30) to be present okay, but in general we can do either of these approaches. The point I showed you this is also because see we do not have to, just because this there are these deep learning tools available to you, you can use that to solve every problem that is true but then many of the conventional techniques are out there, well studied, well-documented and nicely programmed, you can use them actually do make the problems much simpler.

So what we have done by doing this cropping is to make the problem very simple, in the sense that CNN now does not actually have to localise the heart right. We know we are only giving it the beating heart as input, the other parameters are the ones that we try to edges, et cetera would be the ones to focus on okay. Again yes of course you do not want to do any of this that is also fine, you can take the entire image, give that as input to the CNN. Also recall that you know when we did this memory issues were also there but now I know there are more powerful GPU parts and more memory available in them and RAM available, so in fact you can fit very large images and feed them as input to CNN, of course training them is also much harder, I have to figure out the correct hyper parameters for the same ok.

(Refer Slide Time: 20:45)



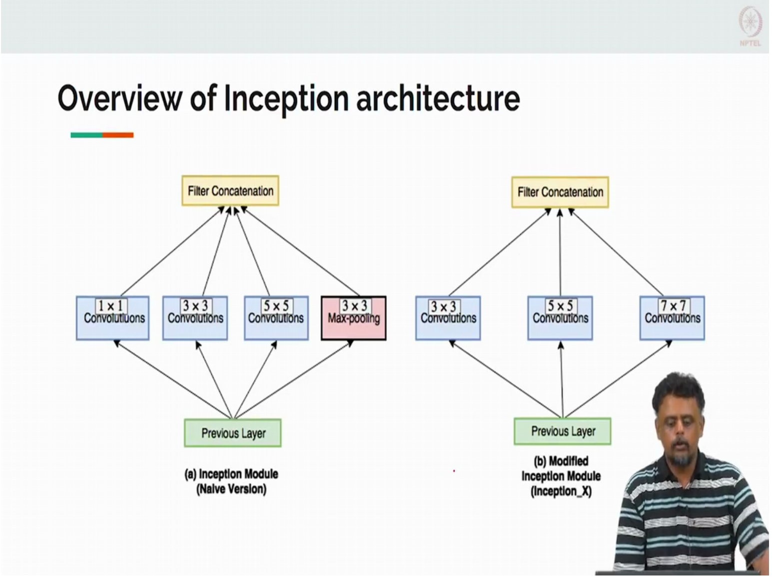
Typical Convolutional Neural Network (CNN)

So what are we going to use? We are going to be using CNN, I know surprise is there because it is a powerful tool for image processing. So once we have just walked you through some of the ones we have already seen, so this is a typical CNN that we talked about where you have an input image or an RGB input image in this case you know you can also have a series of a stack of images as input that is also possible now, we have looked at TD CNN set, I think I mentioned this when I talked about brain tumour segmentation but in this case I think we are going to do plane by plane so this is the input image, then we have a bunch of feature maps extracted by filters and then come to fully connected layer and then the corresponding output categories okay.

So this is the one typical CNN that you have all seen, just to refresh your memory so in this case also remember the problem that we want to solve is the one corresponding to semantic segmentation, so we have to figure out the class that every pixel belongs to, namely the green which is the myocardium or the muscles of the heart especially around the left ventricle, the left ventricular cavity LV cavity, this is myocardium, and the right ventricle okay these are the things that we want to label okay, so they carry the most agnostic information ok. So how we do that, we saw that you know by shifting to our fully convolutional neural network CNN structure then we can do semantic segmentation in one stroke, basically entire image through the one pass all the pixels are predicted okay.

So that is accomplished by you know this is the encoding layer you can think of and this is the decoding layer, decoding layer uses binary interpolation or transposed convolutions to increase the size of your feature maps so that you are able to predict all the labels a one-shot.

(Refer Slide Time: 22:40)



Overview of Inception architecture

Filter Concatenation

1 x 1 Convolutions

3 x 3 Convolutions

5 x 5 Convolutions

3 x 3 Max-pooling

Previous Layer

(a) Inception Module (Naive Version)

Filter Concatenation

3 x 3 Convolutions

5 x 5 Convolutions

7 x 7 Convolutions

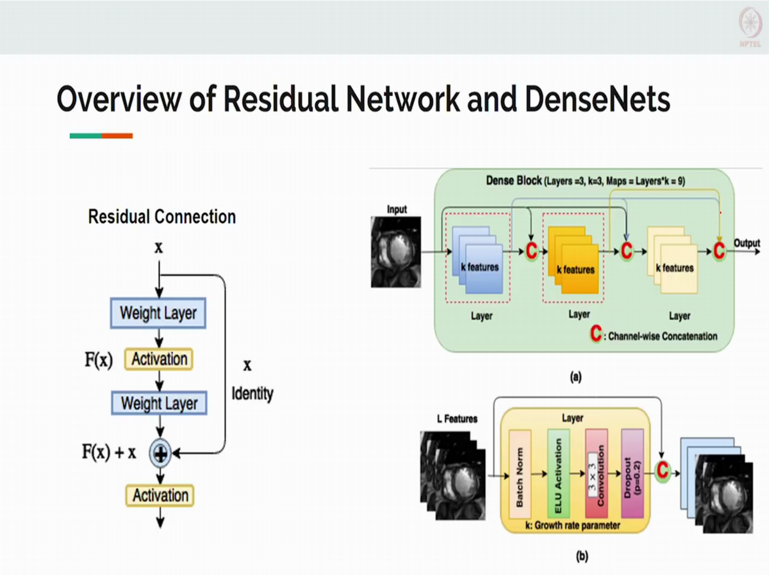
Previous Layer

(b) Modified Inception Module (Inception_X)

And also the encryption architecture wherein we are looking at receptive fields of different sizes in the same layer, so typically in most of the networks we have seen inception architecture, we fixed the side of the field in every layer here, every layer has a combination of different filter sizes based on the problem, it is 1 cross 1 to 5 cross 5 that is the naive implementation. Of course in order to make, in order to make the computation more feasible we also have the possession where we do the 1 cross 1 convolution to reduce the size of each of the feature maps corresponding to different filter caviar sizes ok. So these are the 2 examples here for instance, one here this is 1 cross 1 to 3 cross 3 and 5 cross 5, this one has a slightly larger filter kernels.

Both of them are naive interpretation except that this one what is missing here is the 1 cross 1 convolution, which protect them into lower dimensions right and then you can do the 3 cross 3 ok.

(Refer Slide Time: 23:49)



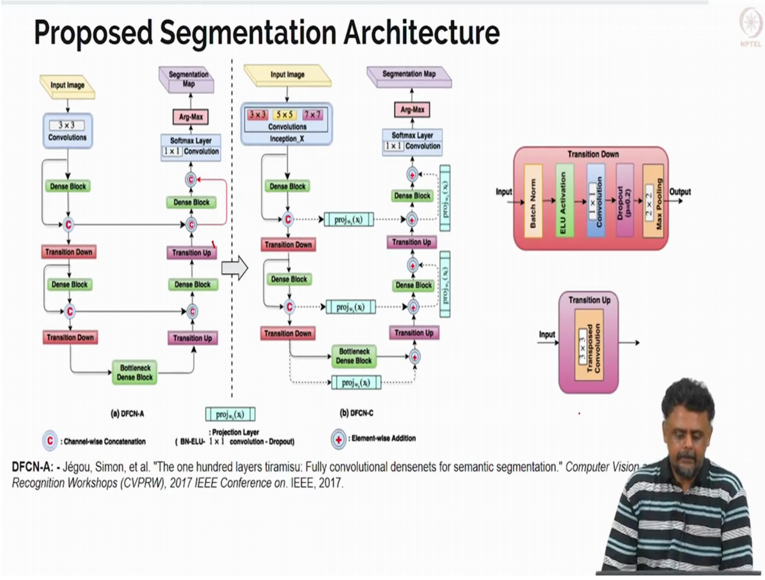
Overview of Residual Network and DenseNets

So the other two concepts again that we are going to be using, the residual network or the dense net, so the residual network we saw that it improves your training faster conversion and you can go deeper by actually moving or adding feature maps from the previous layer into the successive layer. Let us say the input to the previous layer X is then taken as the output rather to the output of the next layer ok so the skipped connections you can do or shortcuts whichever we are going to call it ok, we call it shortcut right, this is shortcut, we call this shortcut connections ok.

So here we add them, so here it is addition so (())(24:36) at the feature map from 2 layers below to the output, so that helps in faster conversion, better learning and makes you do deeper networks. Another version of that is the denseNet ok, here they go, they define a dense block which has several convolutional layers, and one of the hyper parameters to the dense block is called the growth factor or K = 3. In this case what happens is, every convolutional layer inside the dense net block output exactly K feature maps ok. And the output of the block is again concatenated to the input that is what is shown here, so you have L features as input, it goes through a dense block which has multiple layers including convolutional layers and the output of the dense block is: concatenated to the input, so this will be like K + L features okay.

So layer typically one convolutional layer inside a dense block, so this is actually the blow-up of a actually one convolutional layer inside the dense block, we have a batch normally in the activation function followed by 3 cross 3 convolution, so prior to the convolution we have all these other operations which are also defined as layers, so that is just one convolutional layer ok, so in dense block we will have multiple such convolutional layers ok, and each of them will output K feature maps and you actually concatenated the input to that particular layer with output and it goes as again input to the next layer okay, so that is the idea behind dense block that we saw earlier.

(Refer Slide Time: 26:20)



Proposed Segmentation Architecture

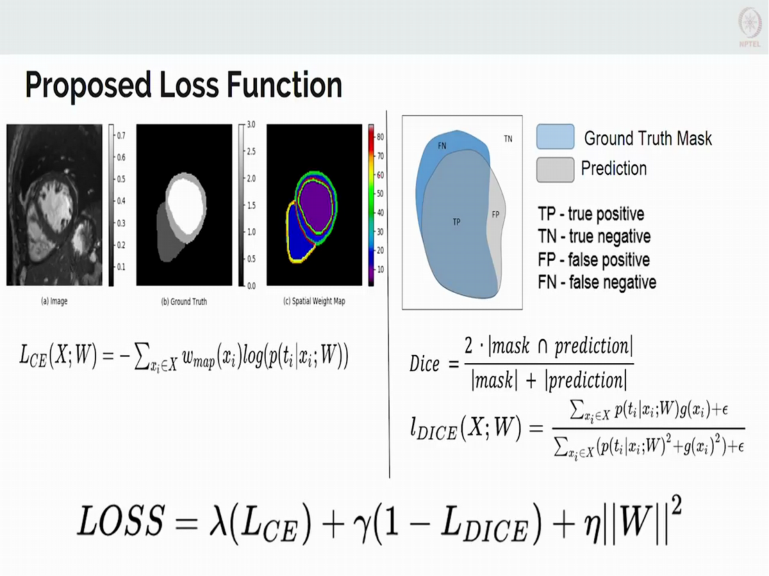
DFCN-A: - Jégou, Simon, et al. "The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation." *Computer Vision a... Recognition Workshops (CVPRW), 2017 IEEE Conference on*. IEEE, 2017.

So the architecture that we are going to look at is similar to the architecture which is reported in the picture is called the, the paper is titled (())(26:27) okay so it has about 100 convolutional layers. So it has the kind of like a similar to the unit, but instead of just playing convolutions you have dense blocks okay. So the transition down the blocks are again, so if you look at it so there is an input where you do a bunch of 3 cross 3 convolution which and output of that goes to the dense block and again the output of the dense block is concatenated, where C is the concatenation with the with the input.

And there is a transition down layer which is again another dense block except that that has that is the one that is blown up here, it has actually max pooling and 1 cross 1 convolution ok, just to reduce the size of the feature map, et cetera. So after a couple of transition base, you have a bottleneck layer again 1 cross 1 convolution and then you do the transition of layers right, so it is the 3 cross 3 transposed convolution. So similar to unit structure, when you go all the way to the top, when you are actually sampling in your which is the decoding side, you also have these connections which will just concatenate because this has the resolution intact, so when you are upsampling it is good to have the resolution back.

So this is the architecture that was reported in literature, so we have made several changes to it, one of them very obvious change is the projection here so instead of directly concatenating, we just project it using 1 cross 1 convolution and we just add it to the layer on the encoding side okay sorry, from encoding side we add it to the decoding side if you think of it that way okay. So these are some of the changes so the idea is it is a unique kind of architecture except that instead of the plane convolution it has dense blocks and we also have these connections going from the encoding side to the decoding side, then there is a direct concatenation here we project and then we add okay.

(Refer Slide Time: 28:46)



Proposed Loss Function

🟦 Ground Truth Mask
⬜ Prediction

TP - true positive
TN - true negative
FP - false positive
FN - false negative

$$L_{CE}(X;W) = -\sum_{x_i \in X} w_{map}(x_i) log(p(t_i|x_i;W))$$

$$Dice = \frac{2 \cdot |mask \cap prediction|}{|mask| + |prediction|}$$

$$l_{DICE}(X;W) = \frac{\sum_{x_i \in X} p(t_i|x_i;W) g(x_i) + \epsilon}{\sum_{x_i \in X} (p(t_i|x_i;W)^2 + g(x_i)^2) + \epsilon}$$

$$LOSS = \lambda(L_{CE}) + \gamma(1 - L_{DICE}) + \eta||W||^2$$

We also have an inception module at the beginning, if you consider different uh receptive field sizes and different resolution ok. The 2nd aspect of the neural network is the loss function, so we do the usual cross entropy loss function so we have about 3 classes. But it is weighted, the weight is obtained from the ground truth so for every image there is a corresponding ground truth here. You can see some of the rarer classes are weighted more than the classes that occur more frequently, so for instance if we look at the red regions here, read regions redline here is the weight map, so this is what you are looking at here is the weight map.

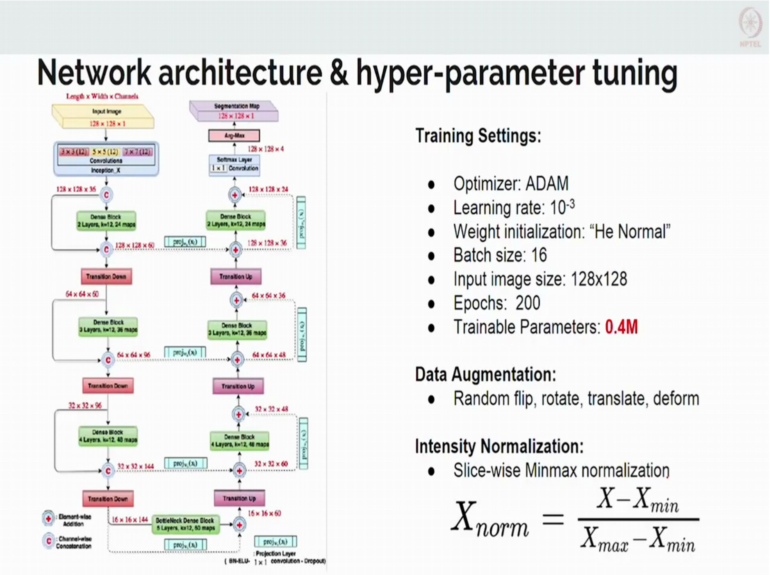
So the red region here has a higher weight corresponding to the other region, and if you look at the blue regions are weighted much lesser, it corresponds to the weight of 20 beats, of course you have normalised this so that it does not blow up, so just to show that in the scale we have shown it this way ok. So this is a rare region because we have a lot of overlap with the right ventricle here so that is difficult to segment, so that edges are weighted more than the other edges. So based on the pixel count, edge location, et cetera, each pixel is weighted differently, so that is the weight map that goes into the loss function okay.

The 2nd aspect of it is that now since we are doing predicting all the pixel labels in one shot, one other way regularising the output is by regularising the loss function by using what is called the dice coefficient. So the dice coefficient what it measures, again this is the formula is the intersection between the predicted mask and the ground truth mask so that corresponds to this region, it is mask truth positive right divided by the cardinality of the mask plus the cardinality of the prediction ok. So you can think of it as the fraction that you got right ok, fraction of pixels that you predicted correctly ok. So but then this if you want to actually use it inside a loss function then you have to make it into a form that is differentiable and that is what is given here, so the P is your output is core from your CNN, so that is the procedure probability it is predicted by CNN and G is the ground truth corresponding to that.

So X i is the pixel label, so every pixel has its label corresponding or probabilistic output and also its ground truth are given, so submission over that divided by normalising factor so that is the differentiable loss function ok. So when you write the total loss function you have hyper parameter lambda multiplying the Cos entropy loss + another hyper parameter multiplying 1 minus L dice, so we want to minimise the loss function that is done by maximising L dice. So if you think about it if you want perfect position then there should be an expect match between your so loss L dice of 1 think about it corresponds to a perfect match between your prediction and the ground truth, so if you overlap them we do not see any difference between them so that response to one.

So if you want to minimise that, 1 minus L dice means then that table corresponds to the current segmentation right because the highest value of L dice corresponds to better and better segmentations typically, so minimising 1 minus L dice is same as maximising L dice ok. And then we have L 2 regularisation, W here are the weights ok.

(Refer Slide Time: 32:15)



Network architecture & hyper-parameter tuning

Training Settings:

- Optimizer: ADAM
- Learning rate: 10^{-3}
- Weight initialization: "He Normal"
- Batch size: 16
- Input image size: 128x128
- Epochs: 200
- Trainable Parameters: **0.4M**

Data Augmentation:

- Random flip, rotate, translate, deform

Intensity Normalization:

- Slice-wise Minmax normalization

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

So now we will look at the architecture in much more detail and also look at you know how it is trained at, et cetera and interpretation of the result so on and so forth, Mahindra will explain the rest of the architecture from here. So what you see in the slide is the network architecture in slightly more detail, so the input is basically a 128 cross 128 image that will crop in, crop out the region corresponding to the cardiac anatomy and given that as input and then we have the inception module here right, which outputs about 128 cross 128 cross 36 after which you know there is 3 cross 3 convolution output 12, 5 cross 5 convolution output 12 and 7 cross 7 convolution output 12 feature maps, these are given as input to the dense block with K = 12 and it gives you about 24 feature maps L output so then you have about 128 cross 128 cross 60 ok.

The transition down layer reduces it to 64 cross 64 cross 60 so we can do that again alternating between the dense block and the transition down block all the way to the bottom here when we get a 16 cross 16 maps 144 of them. So we have a bottleneck layer which outputs about 60 maps and again this is taken by the transition up layer, which goes again all the way to the output segmentation map of size 128 cross 128 cross 1, and in between we have all these shortcuts we call them, which basically take the feature maps of the corresponding level. So remember, as we come down this side, this is the encoding side, the resolution starts to go down and on this side we are trying to up sample so it will be…

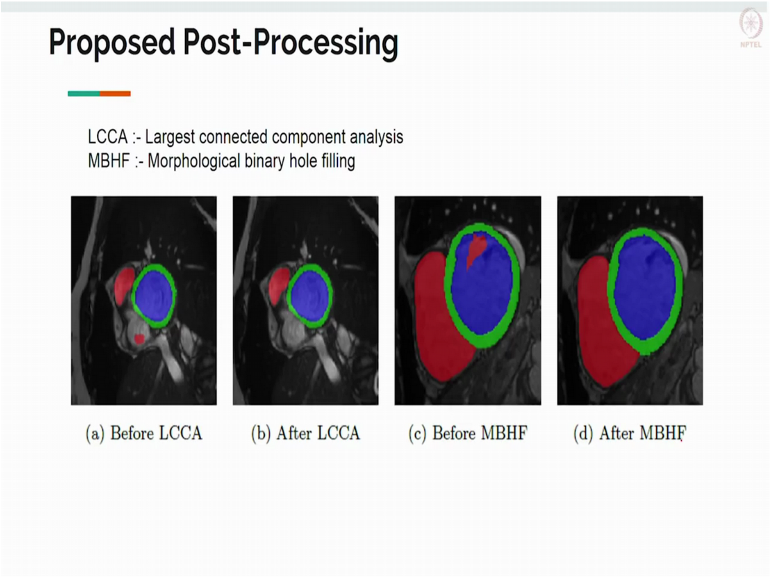
To improve the up sampling we take the information from the encoding side and pass it to the decoding side and that is done by using a projection layer, so we can project the 128 cross

128 cross 60 into one feature maps and add it to the feature maps that come out of the transition blocks that what happens in every layer typically okay. So the optimiser use ADAM with a learning rate of 10 raise to -3, weight initialisation “He normal” weight initialisation is used. Uses the batch size of 16, it is a mini batch size of 16 and it is trained for about 200 epochs and it has about 400,000 parameters.

We also have augmented data by random flips, flipping is you flip the image left side or right side, rotations, translation and some elastic deformations which are very small deformations because you know cardiac anatomy we cannot randomly deform, these are deformations which we actually manually have to look at the range in which we can do. Denser flow, this is programme in denser flow and denser flow has its own inbuilt routine for deformation which has parameters in tune for deformation that are not too that will actually what you call make it you know make it unfeasible, so you cannot randomly deform the heart, so you know there is an anatomy we are looking at so within the confines of the problem that we will be able to produce deformations okay.

Intensity normalisation we use slice wise minmax normalisation that is for every slice it is minimise between 0 and 1, the range is confined to 0 to 1 ok that is the way of doing it.

(Refer Slide Time: 35:47)



Proposed Post-Processing

LCCA :- Largest connected component analysis
MBHF :- Morphological binary hole filling

(a) Before LCCA
(b) After LCCA
(c) Before MBHF
(d) After MBHF

After prediction we do postprocessing, so what does postprocessing to is, does largest connected component analysis. So in Matlab these are automated, so please in Matlab please look up connected component analysis okay. It comes under the topic of morphological image processing, you can look that up to. So we do largest connected component and also hole filling, again morphological binary hole filling.

So what do you mean by that is, if you look at it, there is a nice prediction here but this is not correct so if we do this connected component analysis, one we does is identify the large connected components and we can play a threshold on it, this can be slightly arbitrary (()) (36:19) problem and remove the small one corresponding to every class ok. So in every class we do connected component analysis and remove the ones that are not meaningful right. So we expect the left ventricle to be in one piece and we take the largest piece possible okay think of it like that and so this is a post connected component analysis. We also do hole filling, so for instance this class this is again wrong classification so we have to fill that up with the correct classification that is also accomplished using morphological hole filling. Again Matlab has all of these enabled so you can actually look that up also.

I think Python and other, Python also has some of these morphological image processing tools also available.