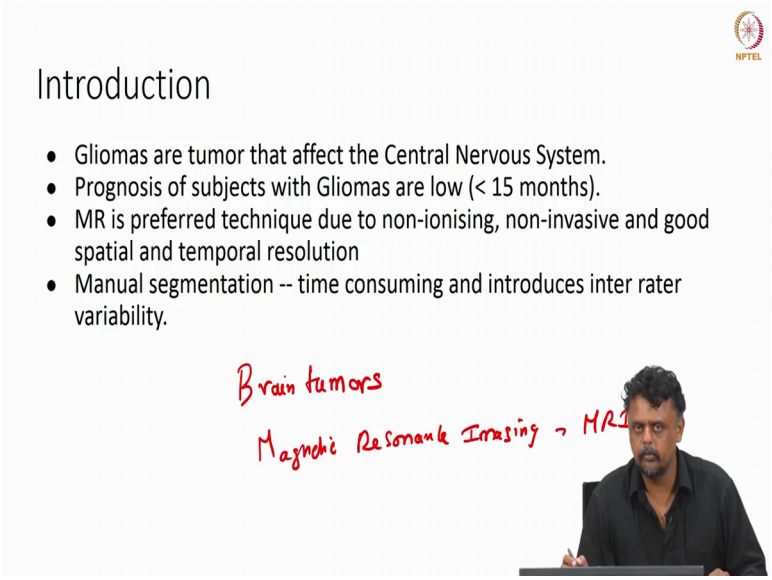
Machine Learning for Engineering and Science Applications
Professor Dr Ganapathy Krishnamurthi
Department of Engineering Design
Indian Institute of Technology, Madras
Segmentation of Brain Tumours from MRI Using Deep Learning

(Refer Slide Time: 0:28)



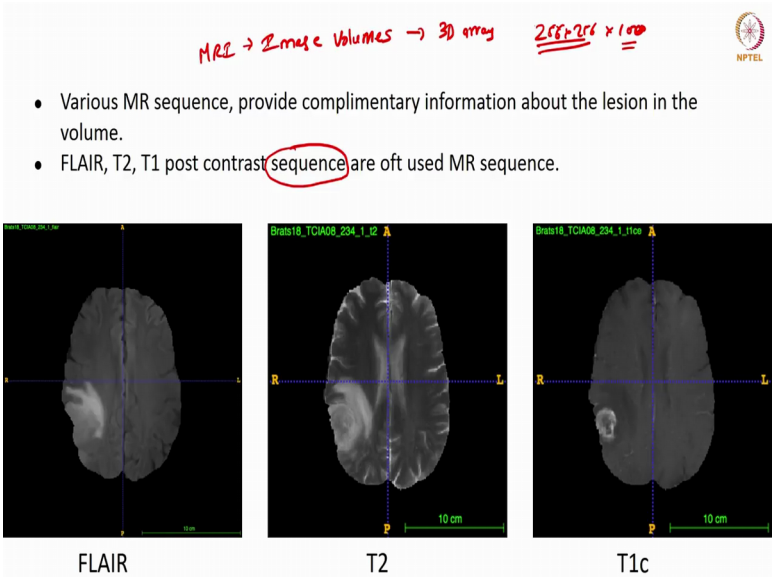
Hello and welcome back. In this video we will look at segmentation of brain tumours from Magnetic Resonance Images, using some of the deep learning techniques, especially CNN's that we have seen so far. So, brain tumours are Gliomas, so we typically refer to them as brain tumours. They affect the central nervous system or they usually are in the brain. And this is a serious illness, form of cancer, which has very poor prognosis, in the sense that the survival is less than 2 years.

And the idea, the patients are typically monitored by magnetic resonance imaging, also referred to as MRI, Magnetic Resonance or MR imaging MRI as it is known as. It is a I call imaging technique most imaging techniques, it is non-invasive, non-ionising radiation is used. Basically a magnetic field and you have RF excitation, that is what enables the meeting. It has got very good spatial interpose, submillimetre's spatial resolution, in this case temporal resolution in this case is not very important but spatial resolution is good.

So the idea behind imaging this patient is that by imaging them you can visualize the tumour noninvasively and by looking at the tumour, measuring its size, one can, Doctor use that as a marker for figuring out if the tumour is progressing result is responding to medication. So the segmentation or delineating the pixels corresponding to the tumour is an important task,

diagnostic task that way. It is typically done manually by an expert radiologist, however it can be very time-consuming and there is some very variability among radiologists on certain tasks.

(Refer Slide Time: 2:44)



And if you have very large patient, you want to do like meta-analysis or anything, then manual delineation is exactly not possible. So in order to augment a radiologists effort, it is the deep learning program that can effectively segment the gliomas, can be very valuable to me. So, typically not one image is acquired but volumes are acquired. So typically MR image volumes, it is called MRI image volumes are acquired. We will not go into the details of the acquisition of the physics behind the acquisition. It is the image volumes, when I say image volumes, they are basically 3-D arrays.

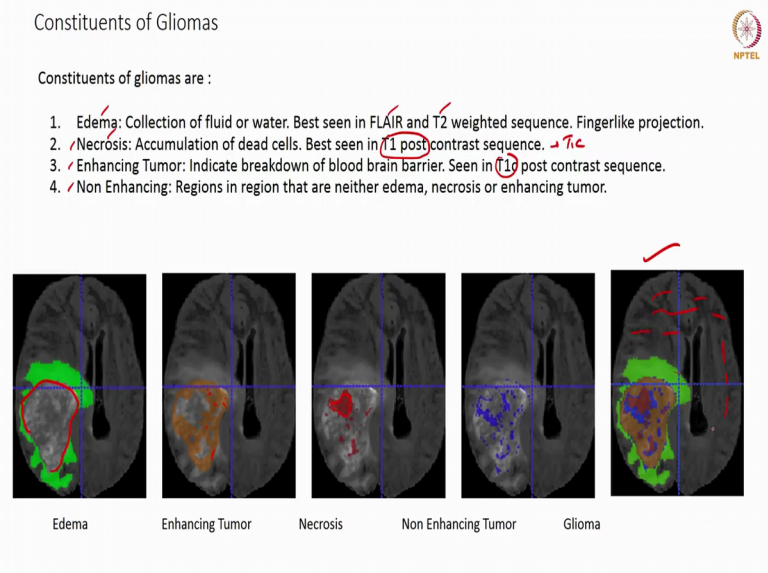
So, each image is a 3-D array. So typically you would say 256 x 256 x 100, okay. So this is the in-plane size and this is the depth, okay, so it acquires the, across the anatomy, okay. So and it is in the form of slices, okay. So the intermittent, so as far as MR is concerned, MR imaging is concerned, multiple image volumes are acquired and each image volume corresponds to what is known as the sequence. This is a, each sequence corresponds to a certain technique or a certain way of exciting the spins inside the human body, exciting the magnetic spins inside the human body.

So, that each of them gives rise to a separate kind of greyscale contrast in the image. So if you look here, we have shown 3 types of MR images, corresponding to the same anatomy. I am just calling to the anatomy automatically and you can see that each of them, we do not

worry about the meaning of the thing right now but you can see that each of these images, even after they pass through the same anatomy, they have a different greyscale contrast, okay. So multiple different types of contrasts are possible using MR images.

So for a typical glioma imaging session, you will typically acquire about for such sequences and each sequence will have a size of 256 or 256 times 100, where 100 is the number of slices through the anatomy. If you are wondering where that 100 terms, so let us say this is a typical human head right here, okay. So brain is right here somewhere. So you would acquire slices which cut through the brain, okay, or the head so that you cover the entire brain or the entire head in this case.

(Refer Slide Time: 5:26)



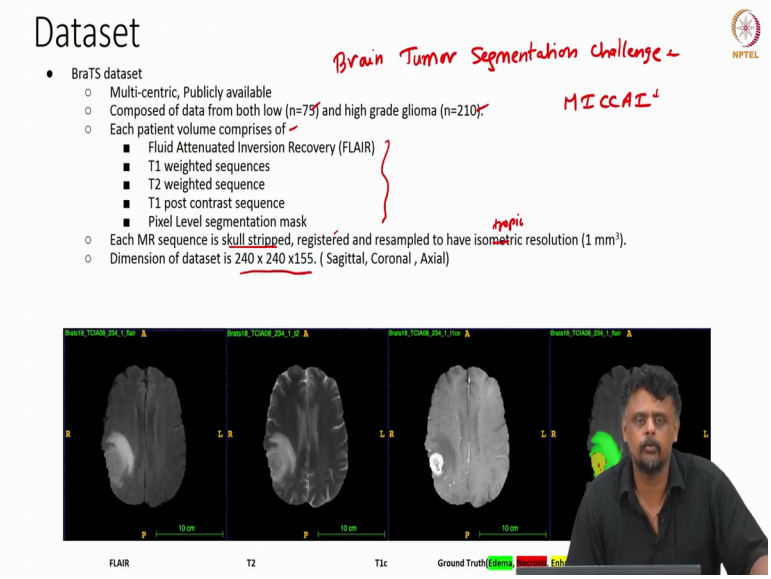
So just to reiterate, every MR image is actually a volume, it is a 3-D array and you will acquire about typically about 4 such 3-D arrays per patient for diagnosing gliomas. So, typically a glioma this is in-line with some conventions that have been laid out. It is diverted into these 4 sub compartments, there is a oedema, Necrosis, enhancing tumour and non-enhancing tumour. We see the image below, if you see in this image, the tumour is delineated this way, kind of, all right.

The green regions are basically what is known as oedema which is some fluid or water accumulation, okay. And this also tells you why we need 4 different sequences because certain components of the tumour are seen much more clearly in certain sequences. So for oedema is seen very well in FLAIR and T2, the necrotic region where they accumulate, the cells accumulate is seen very well in T1 post, start or T1 C it is always effort to as T1 C.

Enhancing tumour, also which indicates breakdown of blood brain barrier is seen in T1 C, again seen in T1 C, so if we look here, the enhancing tumour is, this region is marked in colour, okay.

The non-, the necrotic regions are once again marked by different colors here in another sequence. Okay. Non-enhancing regions are those regions which are major of the 3. Okay, and again there is some variability among regular rests at what these regions are. So the final segmentation, akin to semantic segmentation, that we are looking for is given by this image you see, you read about 4 classes of pixels that we want to delineate. Of course what is left out here is a normal, so everywhere else here, all the pixels here are normal pictures that correspond to different class completely.

(Refer Slide Time: 7:19)



Okay, the classes within the pixels are, the oedema, enhancing tumour, necrosis and non-enhancing tumour. So this is the task. So what does the data look like? So, this data is part of the brain tumour segmentation challenge, which is conducted every year as part of the medical imaging conference call MICCAI, conducted in the different cities in the world. Very, it is one of the conferences for medical image analysis and this particular challenge has been one of the more popular challenges, you get a lot of people entering it, trying to win the challenge.

So its acronym is BRATS. So the dataset is publicly available, it is a multicentric dataset, so to elaborate it briefly, it is multicentric because MR imaging is the greyscale values or the contrast that you see in the values and some of the artifacts and shading that you get in the

images, vary from scanner to scanner and from hospital to hospital. So it is important to get data from different scanners or different centres, like different hospitals, so that your network generalises well to some new data from the different hospitals. And there are 2 types of gliomas, low and high grade.

So the high-grade glioma is the more serious condition. Of course typically more slow grade gliomas progress Slightly primer, I am not talking about diagnosis, just to show you, that there are 2 types of them. And the tumours in them to look different. So it is important to understand that there are different delineations required. And so, in the sense that I will say they do look different. If you train a network on high-grade glioma, it is quite possibly not going to work very well on the low-grade glioma, that is what I mean by different.

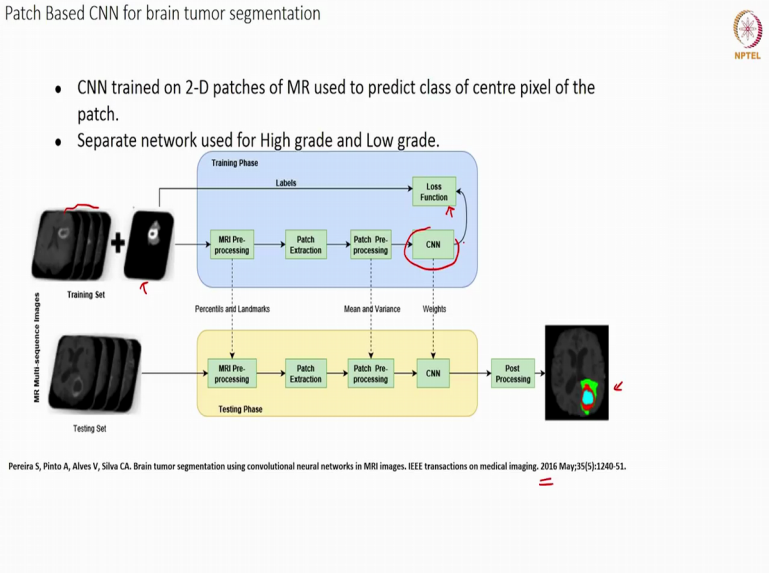
So, if you have to talk to a radiologist to get rarely, if you want to understand the pathological, pathology difference between low-grade and high-grade glioma is anyway. So, each patient volume consists of 4 different volumes are 4 different sequences, okay. So these are all the flair sequence, fluid attenuated inversion recovery, T1, T2, these are the names of the techniques used to acquire the volumes. And if we see on the panel below, you will see that again each technique gives rise to a slightly different or in this case radically different greyscale contrast that helps radiologists identify the pathology in the tumour.

In the image as well as the different types of pathologies. We will not go into these things, which I mentioned here, which is each MR sequence is Skull striped. So, typically the skull is also an image and generally interferes with a lot of processing you would typically try to remove that. They are registered in the sense that not all patients do not have the same head orientation during the acquisition of these different sequences. So there will be slight differences in the orientation, so you correct for the post if you can call it that.

And they are all resampled to have isotropic resolution, in this case it is isotropic resolution, okay. But 1 millimetre cube resolution, okay. That dimension of each dataset, each volume, there are 4 volumes, each of size 240 x 240 x 155, okay. So typically you can slice this along any axis, so since this is a 3-D, you can, we are looking at the 240 x 240 cross-sections typically slice this. And the one 240 x 240 cross-section is what is referred to as the axial. And then the 2 other perpendicular cross-sections are defined, they are referred to as coronal and sagittal.

We will not discuss those further but you can look them up to get a better understanding. So, typically you have these sequence of images per volume and you have 4 sets of volume. And ground roots are given for these datasets, 210+75, the ground truth is corresponding to the colourmap, the segmentations we saw earlier, it is also shown right here. If you see, so each of the tumour class, intact tumour class is marked by different colors. And the task is to obtain a similar classification by using some machine learning techniques or other image processing techniques.

(Refer Slide Time: 12:10)



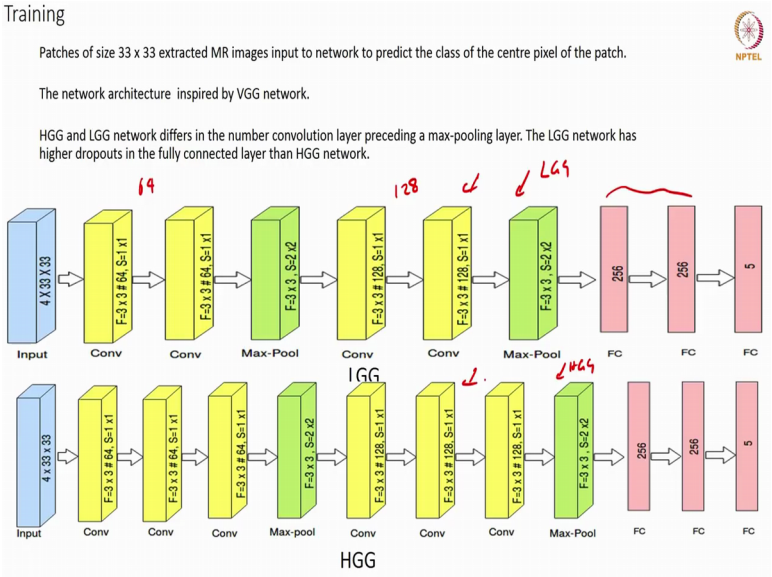
So we will typically explore the CNN's that have been very successful at this task. Okay. So the 1st CNN that will look at corresponding to this publication, this is one of the remaining entries to the I think to the 2015 competition. Here they have used the CNN trained on 2-D patches, so if you recall from our discussion on semantic segmentation, one of the naive ways of doing semantic segmentation is to extract patches from your images and label the centre patch according to the image class.

So that is the typical strategy followed here too. So you have a training set here 4 sequences, okay and you extract a patch, centred over a pixel. So you will expect the price from all 4 sequences. So, you put these 4 channels with a patch corresponding to each of the 4 sequences. And you will only look at 2-D patches, not 3-D patches, we will look at really patches later. And of course the ground truth corresponding to that, we have it and prior to extracting the patches, there is a lot of pre-processing done, because these MR images are from multiple centres.

So there is a histogram matching step just to make sure that the intensities in the images, so if you consider the intensity of 100 or some anatomy in the brain, which has an intensity 100, you want that to be hunted in all the images. So just to do that, you do some histogram matching to match the distributions of the pixel values are across the image volume, across the dataset. So, there is a patch extraction and pre-processing steps done and you train it with a CNN correspondent to a loss function.

So this is a classification task, appropriate classification is used, okay. And then when you went for during the testing phase, you use the trained CNN here into get your final output labelled. Okay. So, here the labels, there are one of 5 labels, right, you would have background, label, or, and you have the 4 classes inside the tumour, okay.

(Refer Slide Time: 14:38)



So, the network used is inspired by this BGG network. We have a succession of convolution layers followed by Max pooling and as they go, so these 2 layers of 64 filters defined, these layers of 128 filters defined, okay. And you have a Max pooling layer of 2, and then you are fully connected layers and you have output, here 5 class output, okay. So, this particular group trained 2 networks, one for high-grade glioma and the other one for the low-grade glioma. So it is actually another way around. This is the low-grade glioma network and this is the high-grade glioma network.

You can see here is the high-grade glioma network has slightly more number of convolutions, this case 3, convolution of 128 feature maps defined. And there are Max pooling layers are supposed here, the number of convolution layers here is usually 2, it has 3 and it has 2. So,

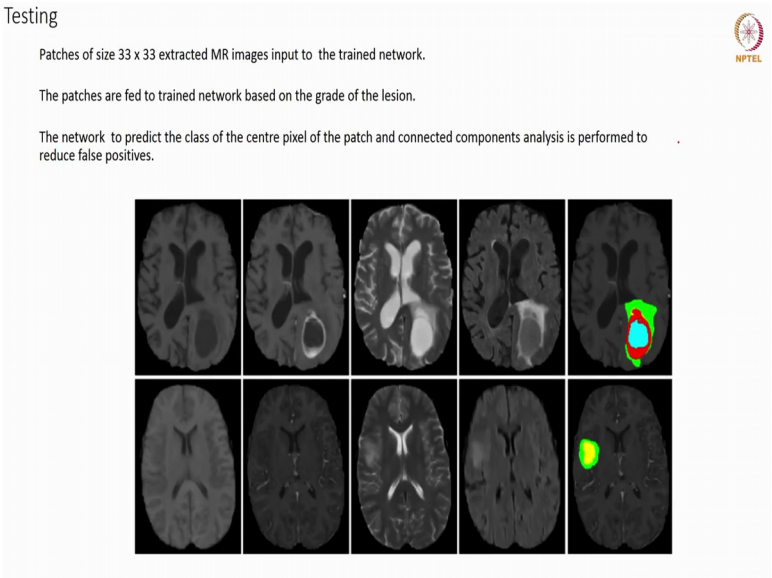
there are 2 successive convolutions with 128 feature maps with 3 of them here. So they train networks with data, corresponding to low-grade and high-grade with these 2 networks. And they have the most competitive performance for the segmentation challenge, which was in 2015 is what I recall.

This is inspired by VGG, but the number of filters and the number of layers is much different, okay. You can also understand that you cannot have a network as deep as the VGG network because do not have the point, data points. And we are training, the training was done using the patch that we saw earlier, using patches. So you can actually typically extract, since the size of your input volume is 240 x 240 x 155. And you typically, this group extracted patches of size 33 x 33. Of course there are 4 sequences, for your input has 4 channels, with patches of size 33 extracted from each of the sequences.

You them together, that is your input. The output is to classify the centre pixel, in that patch as belonging to one of the 5 categories that we saw earlier, okay. So the 5 categories being it is a normal tissue, enhancing tumour, non-enhancing tumour, necrotic core and oedema. So, these 5 classes, these outputs and it is inspired, that was inspired by VGG. This is one of the earliest CNN reviews for this particular task, That won the challenge again.

And as you can see, some takeaways here being, use the architecture, that is used in image net, it is inspired by the image that VGG architecture. And it is patch based training, so as to label the centre pixel. And of course this entire network was trained from scratch, no weights were used from the VGG network and you also, they also have to do data augmentation by flipping, rotations and translations.

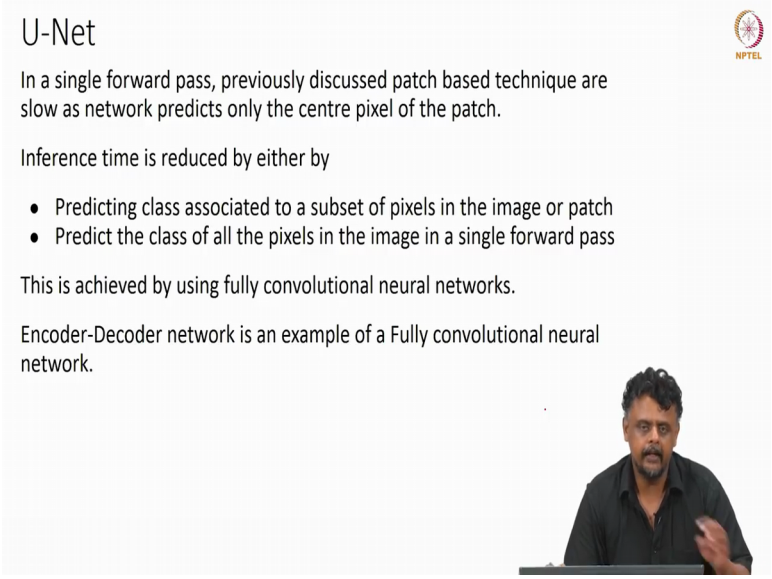
(Refer Slide Time: 18:13)



Just to summarise, once again the idea is to use patches of size 33 by 33, and are 2 different networks, one for the lower grade as the higher grade glioma. And then it was a strange to the network is trained to classify the centre pixel of the patch. And following that, there is something called connected component analysis, wherein you retain the largest connected component of pixels. Connectivity is defined as 4 or 8 connectivity, so if you have a grid, let us say VGG on 3 by 3 grid covers all the ifs, labels belonging to A class.

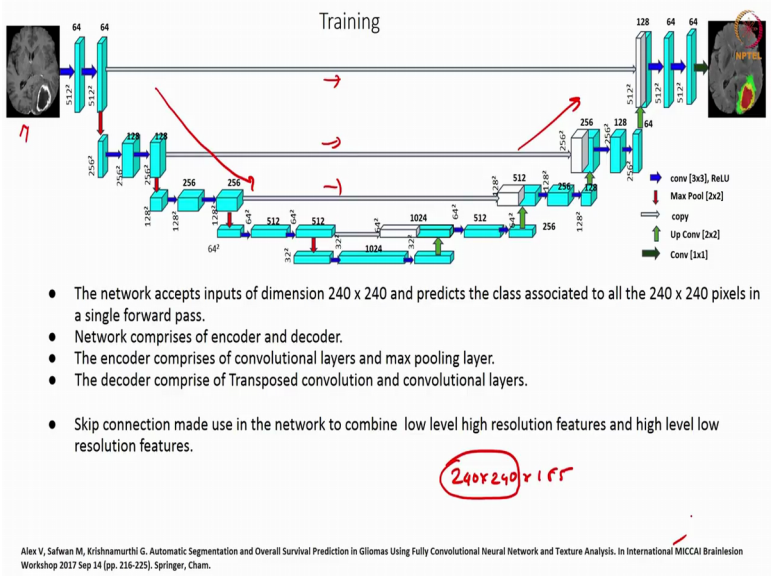
Right, so they are all connected, so you retain them. Let us say somewhere else in the image you have one isolated pixel and you typically tend to ignore it, okay. So you would group the pixels depending on how they are all connected and then remove those pictures that are, that are not corrected in the sense that you will find the largest connected component this way, based on this kind of connectivity. And actually remove those other groups, even though they are connected, they are much smaller than this, so they are removed. So by retaining the largest connected component, they were able to get a very good cause.

(Refer Slide Time: 19:50)



these are the typical segmentation that you get from the network. Another approach to this problem, is using a little U net, U net architecture we saw earlier. So here we can actually predict an entire slice. You try to predict an entire slice and entire patch in 1 task. Rather than in the previous technique that we saw, only the central pixel was appropriately classified as belonging to one of the 5 classes. So, we can use a fully convolution neural network. So that was one of the entries in the competition. Or the encoder decoder type network, which is a U net type, can be used to accomplish this task.

(Refer Slide Time: 20:21)



So let us see how that works. So, here you can give the entire slice as input. So we look that the, again, once again the size is 240 cross 240 and there are about 155 such slices, this is one

slice. So the previous network that we look that, the each slice is last raised, I mean, you have taken a patch of size 33 by 33 you centre that patch on every pixel and try to predict every centre pixel of every patch. That takes a lot of time, so you can use surf using fully convolutional neural network predict the labels of the pixels in one forward pass through the network. Okay.

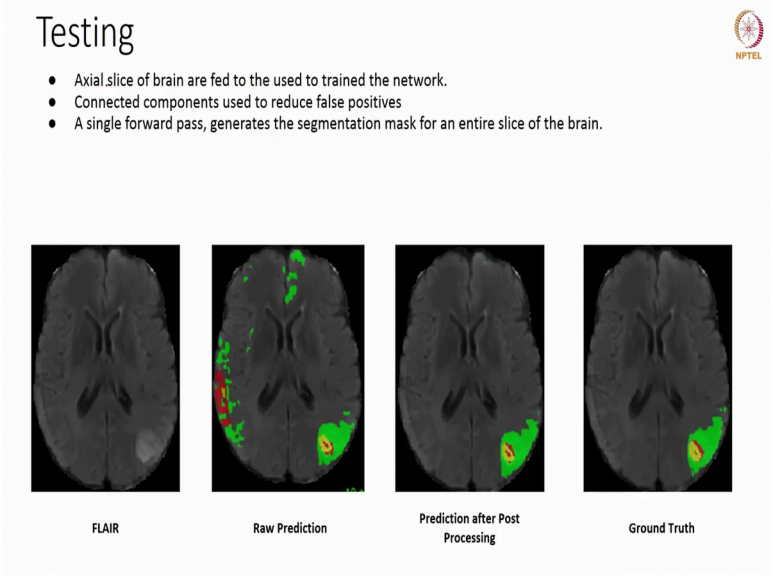
So this is a typical unit architecture that we have seen earlier. Okay. So it takes as input 240 cross 240 and predicts the classes associated with it with all of the pixels. So it has encoder and decoder type of architecture, this is seen through before, so this is the downsampling layer and this is upsampling layers and there are these, we can call them shortcut connections or skip connections from the downsampling layer to the upsampling layer in order to improve your resolution. Okay. So, get the script connections were used to come by the low and high level features in a network.

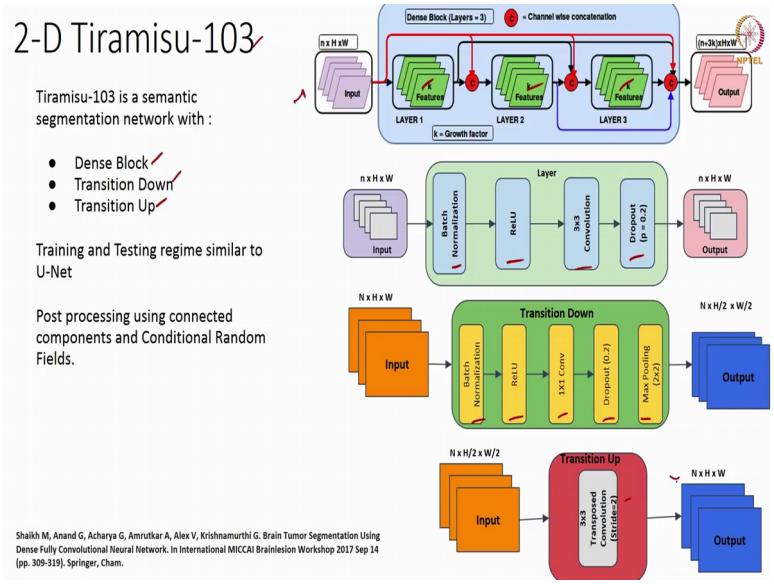
So this was one of the entries in the competition and the citation is at the bottom of the page, you can look them up for more information. Of course here, this might not be done or stay incredibly efficient way of doing it, the better method would have been to predict, let us say a smaller area within the network, okay. Because as you see there is not enough information, there are 2 things, 2 problems here, if you notice, we have mentioned earlier the images are 3-D, so these are not two-dimensional images, these are three-dimensional images.

And it is more meaningful to process them as such. So the slices, successive slices as you go through the volume are correlated. So, it is good to exploit that correlation. And within a slight, so if you look at one cross-section, one slice which is shown right here, there is not much information about this class, the edges of the image, you are losing out because there the neighbourhood information is missing because you are close to the edge of the image. So it would be more meaningful just to predict, let us say point out some smaller area inside the box, inside the image, even using a unit, that is more meaningful.

So you can always take a crop and combine and use that in the skip connections, rather than using the entire feature, the size of the feature map, that would have been more accurate that way. And another thing to point out, again to point out that we are not considering the volume, the image volume, rather than only looking at the cross sections individually. Okay.

(Refer Slide Time: 23:42)



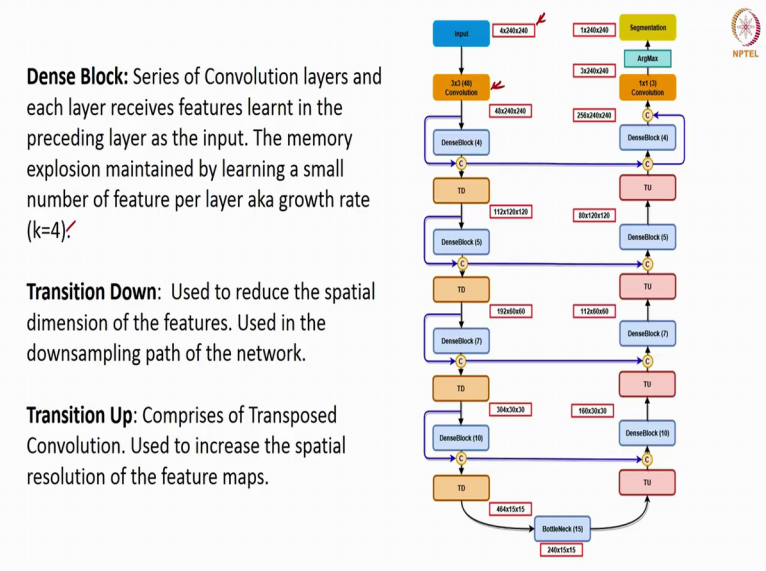


So this is one of the results from that U net segmentation. We have seen quite a few false positives which as I mentioned earlier, if you do connected component analysis, wherein you retain the largest cluster and you can get rid of the smaller ones, so that is a good analysis. On top of that you can also do conditional random fills, okay, but it is not done here. But if you look at the position after post processing, it is typically connected components and compares very well, qualitatively in this case compares very well with the ground truth annotation.

In this video we will look at another architecture for brain tumour segmentation, it is called the 2-D tiramisu with 103 layers. So, basically it is inspired by dense net architecture. It has dense blocks, transition down and transition up blocks. The dense block has 3 layers, 3 convolutional layers and each layer has a composite operations consisting of batch norm, Rel

U, 3 op convolution and the dropout layer. Okay. The transition down layer within your subsampling feature maps, again the batch norms, Rel U, knowledge rarity, again the one cross one convolution again will dropout, layer has been added here during training and followed by Max pooling.

(Refer Slide Time: 25:26)



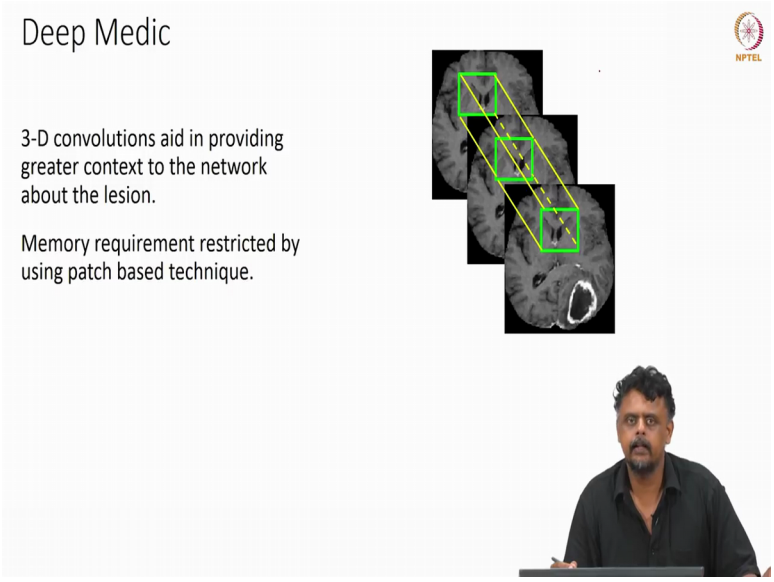
The transition a player has 3 cross 3 transpose convolutions, the slide of 2. And we try to add the network actually predicts the entire input in 1 pass through the network. So, there is block as you seen before, series of convolution layers and each player receives input from all the previous layers. So, just another 2, prevent the feature, feature maps, too many feature maps and it is feature map explosion, we control the growth factor to 4. The transition down layer as we saw before is used to reduce the special dimension of the feature, using the downsampling side in the network.

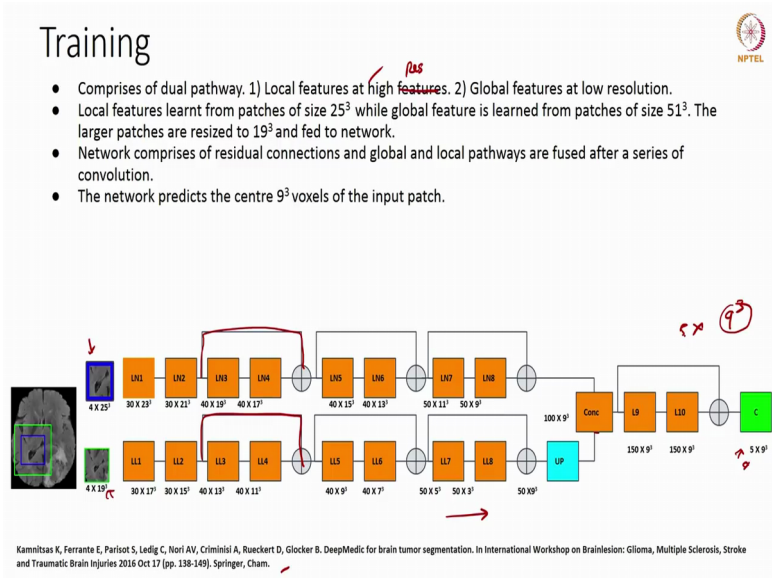
And the transition up is the transpose convolution layer, used in the upsampling side of the network, okay. So the typical architecture is given here, so you have 4 panels as input, corresponding to the 4 different MR sequences, each of size to 240 cross 240. And then of course followed by, there is an initial involution, okay and which leads to 48 feature maps followed by a dense block and a transition down block. Thereafter the transition down block you have the feature map size is reduced to 120 crore 120.

So, you go on, you have several about 1, 2 and 3, 4 dense blocks gives rise to a 15 crore 15 feature maps, 464 of them, you have a bottleneck layer which does T1 +1 transitions and transition up. And then again, we go through the transition up layers which are again 1, 2, 3, 4

dense blocks embedding with intersperse of transition blocks to get an output of which has a 3, in this case we are predicting 3 classes and you can do argument across the classes to get it 240 cross 240 output. The reason why this is done, only 3 classes instead of the 5 we saw earlier because of the classes were merged in this version of the BRAT segmentation challenge, so we ended up predicting 3 classes.

(Refer Slide Time: 27:25)





The other architecture that we are going to look at, it is called Deep Medic from a group in UK. They exploited the 3-D nature of the input data. So they cut slices, all correlated across the volume and of course since your volumes of size 240 cross 240 cross 155, it is not possible to get 4 such volumes as input to the network. You will run out of memory

eventually since you remember that you have to have mini patches and all that during training.

So the idea is to restrict the size of the patches so as to not have the memory issue, but at the same time exploit the 3-D nature of the images. We just briefly look at the architecture used. This was again the winning architecture 2 years in a row, I think 2016-2017. So, this group used 2, citation is at the bottom, this group used multi pathways or 2 pathway network. So one is, one pathway is supposed to give local features at high resolution and the other one has global features at low resolution. So the local features are learned from 3-D cubes of size 25, so, 25 cube and 4 such channels.

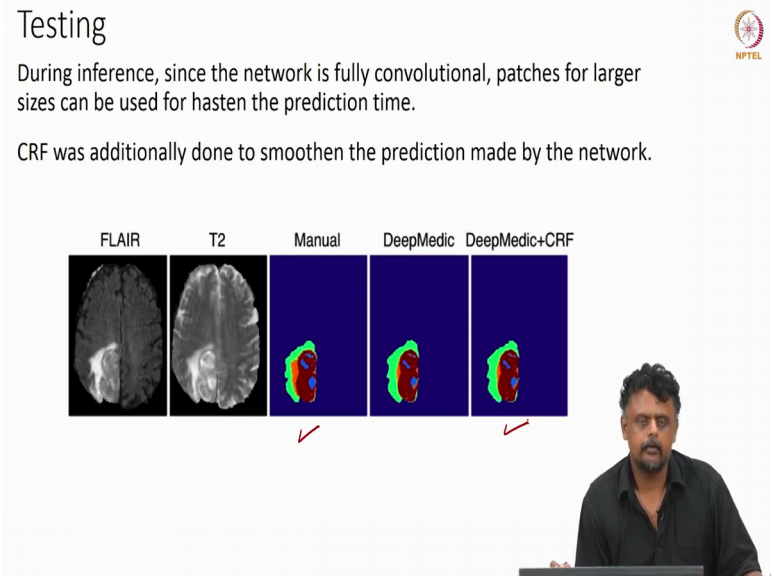
The global features are learned from patches of size 51 but then resized to 19 cube. So, the large green box is a size of the 51 cubes size, 51 size patch, then it is resampled to 19, okay but it is 19 cross 19 cosine volumes from inside the image and there are 4 channels, we have 4 such cubes. And then you have sequence of convolutions and then there are also in between there are these resnet residual layers like in the resnet architecture. And then finally this, the lower the come on the lower pathway which is the global feature pathway is upsampled and concatenated.

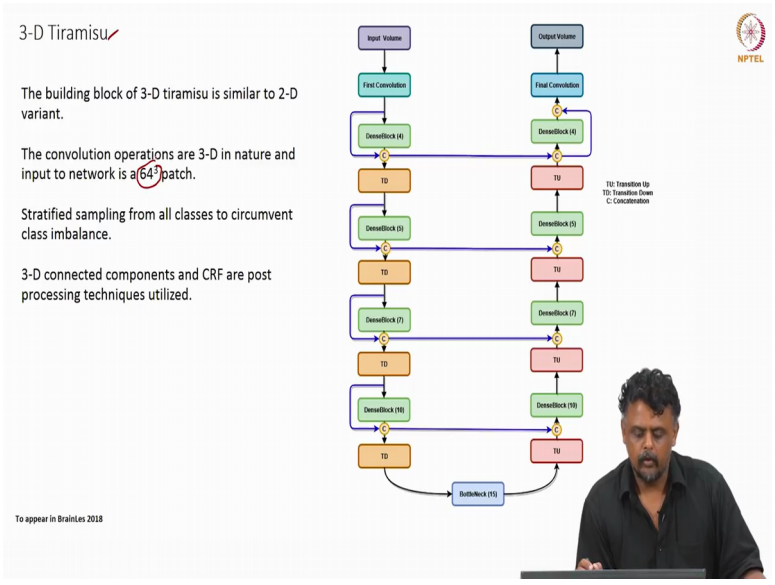
And then you have the usual convolution on fully connected layers which leads to, usual fully convolutional layers which leads to an output of size 9 cubes. So the output size is 9cubes. So you are taking a very large context, which is typically 25 cube and you operating 9 cubes out of it. Of course, you have 5 classes, so you have 5 class, 9 cubes output, each giving the probability of that particular class. So this architecture was very efficient and it won the challenge 2 years in a row. It incorporates several things, one is that we need 3-D context, especially for medical images, that is very important because the images are inherently 3-D, so it is good to exploit that.

It gives us 2 pathways, one pathway looks at a higher resolution but a small patch size. Other pathway looks at a lower resolution or global features but at a bigger patch size. So, you resize the 51 patch to 19 cube and do a sequence of convolutions, Max poolings, as well as the skip residual connection. So, this particular architecture incorporates and of course incorporates both those constants that I mentioned earlier. And also that instead of trying to predict the entire remit in one pass, right, trying to predict only the smaller, of course then, smaller volumes, but then you do have to raster through the network in order to obtain.

So it is not one pass to get the entire volume, but multiple passes because you are only, eventually predicting a 9 cubed. So, you are predicting a small subset of pixels inside your patch that you have taken. But then you are considering information from a very large neighbourhood of the patch. And that is easily handled by the network.

(Refer Slide Time: 31:46)



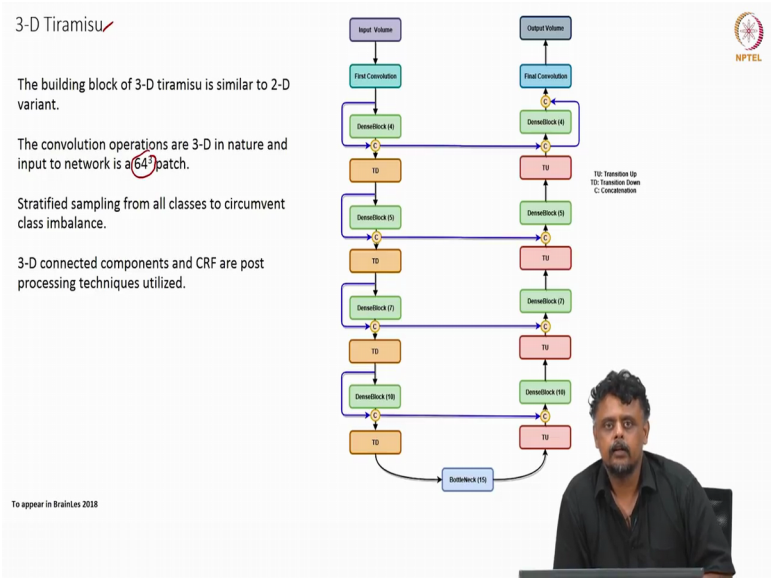


They also have the conditional random feel regulariser at the come after the predictions. So if you look at what these images show, the manual, this is the ground truth and this is the net output predicted by the network. Okay. So the CRFs are used to regularise your output traditions. So that is one of the other normal TV in this particular submission. So a variation of that, not, with assets of using 3-D patches, the 3-D theorem is used in the network. Building blocks are very similar to the 2-D variant, however, patches are 3-D in nature.

Search in of using the entire 2-D volume as input, you take 64 cube patches as input. And you try to predict all of them in one pass through the network. Okay. One of the challenges in this great tumour segmentation is that there is a huge class imbalance, which is finally, as a final talking points in this application that we are looking at. So if we typically look at these images, you will find out that you know some classes are wholly underrepresented. So, for instance the non-enhancing tumour or the necrotic core comprised of a very small number of pixels corresponding to the tumour itself.

Now if we consider the tumour itself, it occupies a very small while in the entire brain. So, less than 5 percent of the tumour in the brain actually corresponds to the tumour, okay. So, that is the class imbalance. So if you are trying to do a pixel wise classification, you see that most pixels are normal, right. Every small percentage of the pixels actually comprise the tumour itself. And within the tumour itself, there are these classes which are very underrepresented. So, these considerations one has to take care of when you train these networks.

(Refer Slide Time: 34:04)



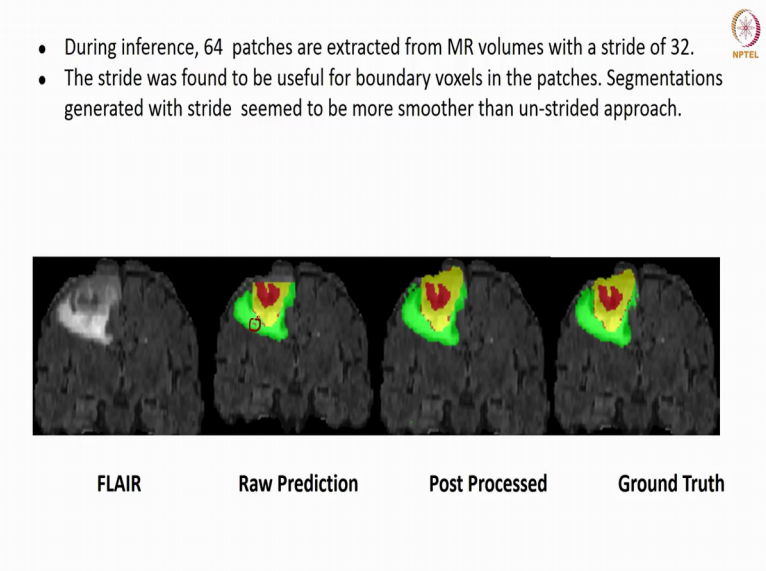
So, for instance, you should make sure that you have cubes at the, cubes your sampling or if you are doing a patch-based approach where you are classifying centre pixel, then you should sample or to motivate augmentation for the underrepresented classes and train the network accordingly, okay. So all the networks you have seen so far using area 3-D patches. We will suffer from it in varying degrees. So the advantage with using 3-D patches is that the class imbalance is slightly elevated because if you are looking at a slightly large volume through

the tumour, then the volume will contain enough samples of every class, that is a general observation that we can make.

So doing 3-D sampling, doing creative patches for medical image and you see in this problem, at least the problem that we are looking at alleviates the class imbalance problem to some extent. Of course you can always say, you can always sample so that you have the underrepresented classes have more samples to match those classes that are overrepresented. And typically we also see that many of the techniques use the collected components to gather the largest connected component or they do a conditional random field approach to clean up the segmentation.

Because the cleaning up requires, in the sense that there will always be false positives, some pixels, some small clusters of pixels or group of pixels being labelled as tumours when they are actually normal. So if you want to remove those, then you can either do a connect component or CRV. CRF takes in the context, so that is a better way of doing it. But most of the research groups a matter not trying to do this postprocessing, rather just use the output from the network itself as a final result.

(Refer Slide Time: 36:06)



So, you can see the impact of the post processing here on this particular segmentation . So, if you look at the postprocessing, so this is a postprocessing image, this particular small region has been removed corresponding to one of the classes and it matches very well with the ground truth right here. So postprocessing helps improve your score a bit. And when you are doing this cube 64 cubes patches, the strident which you sample also seems to help, overlap,

the slide will overlap, especially found to be useful for classification, classifying the boundary walls and the purchase.

So, segmentation will stride, seem to be more smoother than unfettered approach, so that is something that we have observed in our experience. So, postprocessing is of course necessary in many cases depending on how accurate your network is and the amount of false positives that it generates., That is important aspect of the processing pipeline. So that concludes our case study, we were looking at using convolutional neural network for analysing medical images. Specifically to segment brain tumours from MR volumes. Particularly challenging problems since brain tumour is diffused with no clear-cut boundaries in many cases.

The problem also involves segmenting subclasses from the tumour which are again not that well-defined in many cases. The size of the dataset is also a challenge, these are 3, inherently 3-D data and every patient or every data point is actually made up of 3 volumes, sorry for volumes. And how to extract patches from those volumes, whether 3-D or 2-D and train them also to do the influence more efficiently, so that if you get a patient volume, you should be able to do that in a reasonable time, that is also an important requirement, both accuracy and efficiency in competition. So that concludes our session on CNN's, so we will want to other deep learning architectures in the subsequent lectures, thank you.