Image Segmentation of Brain Tumours using Convolutional Neural Networks

Petter Eriksson **Marcus Hägglund** **Fred Lindahl**

Abstract

Segmentation of brain tumours in medical images is an important step when diagnosing, monitoring and treating the disease. This is usually done manually by a radiologist, but recent years have seen an increase in the use of convolutional neural networks to segment tumours in MRI images of patients. In this paper we use two autoencoders, U-Net and LinkNet, to replicate results achieved by other researchers using a similar approach. These networks use a series of encoder and decoder blocks to classify each input pixel as either belonging to a tumour region or not. We achieve DICE scores of 0.703 and 0.715 for U-Net and LinkNet respectively, falling just short of our target of 0.73. Code can be found at `https://github.com/mhagglun/DeepLearning-Project`.

1 Introduction

This report is concerned with brain tumour segmentation on Magnetic Resonance Imaging (MRI) images. Tumours are abnormal lumps of cells formed due to uncontrolled tissue growth [10]. Some tumours are benign, meaning that they do not spread to neighbouring tissue nor to other parts of the body. Tumours that spread are called malignant or cancerous. Even benign tumours can be life-threatening if formed in the brain [6].

A low-grade type brain tumour, like glioma, has an average survival of seven years whereas more aggressive tumours often lead to a significantly more rapid deterioration of health and consequent death [6]. Tumours exert pressure on the surrounding tissue which causes disruption in normal brain function. Brain tumours are difficult to treat surgically as removing too much of the tumour may inadvertently lead to removal of healthy brain cells which can cause severe and permanent brain damage. On the other hand, unless the tumour is removed in its entirety it will continue to spread [3].

To maximise the chance of survival, it is crucial that tumours are detected at an early stage. Extracting the tumour area from an MRI image is critical both for diagnostics and for treatment. However, manual tumour extraction is time-consuming and can be prone to errors [10]. Instead, it is possible that state-of-the-art image segmentation methods may serve as complement and aid to a radiologist performing the task. Early diagnosis is crucial in treatment efficiency for many types of tumours [6].

In this report, we propose deep convolutional neural networks (CNNs) as a means to perform semantic image segmentation. We present two separate network architectures capable of solving that problem. The purpose of this report is by no means to produce new state-of-the-art results, but rather attempting to replicate what has been achieved by other researchers, using identical or similar methods, with our main comparison being an article by Sobhaninia et. al. [10], see Section 2 and their DICE score of 0.73. We implemented the LinkNet and U-Net architectures to train networks that would predict the location of a tumour given an MRI image. We were able to achieve a DICE score of 0.703 on the test set for U-Net and 0.715 for LinkNet.

Project report for DD2424

2 Related Work

This report was mainly inspired by Sobhaninia et al. who used a CNN architecture called LinkNet to segment brain tumours in MRI images taken from three different angles: the sagittal, axial, and coronal planes (see Figure 1). They train three separate networks on images corresponding to each angle and test their networks accordingly. They were able to achieve a DICE score (see Section 4) of 0.79 for a network trained and tested on images of the sagittal view. The network trained on images from all angles achieve a DICE score of 0.73 [10].

The Multimodal Brain Tumor Segmentation Challenge (BraTS) is a yearly competition held since 2012 that aims to assess state-of-the-art models for brain tumour segmentation. All competitors are given the same dataset (which will change from year to year) to train and test their models on. There are three tasks in this: segmenting gliomas in pre-operative scans; predicting the chance of survival for a patient; and quantification of the uncertainty in the segmentation. A winner is selected for each task and presented after the competition. To assess the performance of the image segmentation on the given test set, the DICE score is used again [7]. The most recent winner, with a paper published on the results, is Myronenko who won the segmentation task in 2018 with a DICE score of 0.8839 on the test set. The algorithm used a encoder-decoder structure with a regularisation on the decoder. [8]. Others, such as Havaei et al. who reached third place in BraTS 2013, achieved a DICE score of 0.88 with their best algorithm. They used a novel network structure that exploited local and global features [4]. In general, top performing teams achieve a DICE score of around 0.80-0.88 [11][7].

Although the use of CNNs has increased over the past few years and they outperform other state-of-the-art methods in visual recognition [9], it remains relatively new as a method within the realm of medical imaging [4].

3 Data

The dataset consists of MRI images of 233 patients with brain cancer collected by Nanfang Hospital, Guangzhou, China and the General Hospital, Tianjing Medical University, China during the period 2005-2010 [2] and can be found at `https://figshare.com/articles/brain_tumor_dataset/1512427`. This is the same dataset used by Sobhaninia et al.. There is a total 3064 grey-scale images with varying amounts of images for each patient, although each patient has at least one image from each angle: sagittal, axial and coronal. Additionally, there are usually a few images taken from the same angle but at different "brain slices". The images have an in-plane resolution of 512x512 with pixel size 0.49x0.49 mm^2. The slice thickness is 6 mm and the slice gap is 1 mm. Some images had a deviating size so they were resized in order to facilitate training. Pixel values were normalised to have values between 0 and 1.

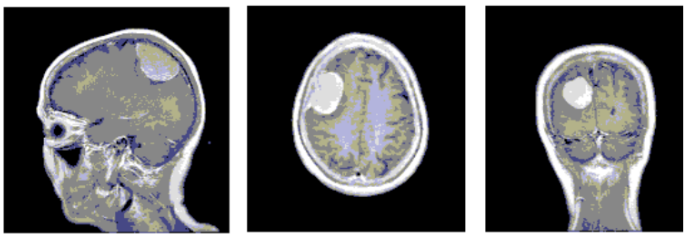
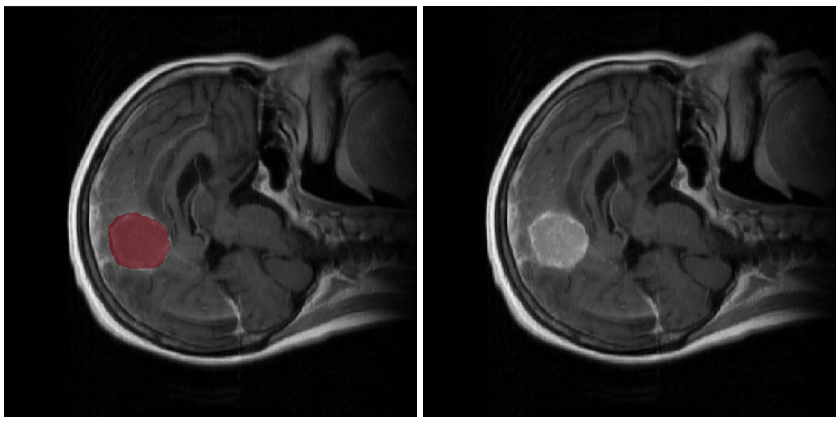


Figure 1: Sample image showing the three views of the brain [10].

Each image has a complementary mask image indicating the location of the tumour, which is used as ground truth values for the network (see Figure 2). The border serving as the basis for the mask was manually delineated by three experienced radiologists.



(a) With mask: red area shows location of tumour.

(b) Without mask.

Figure 2: Side-by-side comparison of an MRI image with a mask and no mask. The large white areas show the tumour.

4 Methods

As we are trying to replicate results achieved by other researchers, we settled for two commonly used network architectures: U-Net and LinkNet (see Section 2). Both of these networks are essentially autoencoders developed specifically for image segmentation. In addition, they are well-documented which facilitates the implementation process.

Image segmentation is fundamentally a classification problem: assigning to each pixel a label indicating whether the pixel belongs to a given mask or not. Both architectures are based on a sequence of *downsampling* or *encoder* blocks followed by the inverse sequence of *upsampling* or *decoding* operations. The purpose of the encoder is to learn a hierarchy of key low-level feature representations. This is achieved by using sequences of convolutional layers followed by non-linear activation functions. In both LinkNet and U-net, convolutional layers with a kernel size of 3 are used to reduce the dimensionality of the input images. The encoder is followed by a decoder that performs a series of *deconvolution* operations in order to restore data dimensionality up to the original input size where the network output is fitted to the corresponding mask image (ground truth). The network architectures are shown in Figures 3 and 4, respectively.

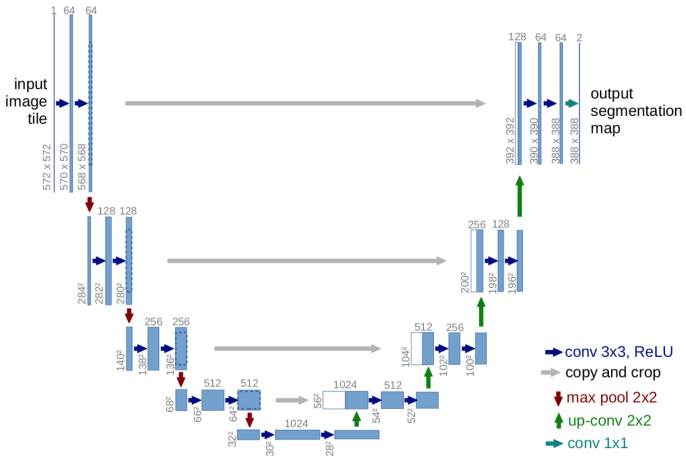


Figure 3: Example U-Net architecture, image taken from [9].

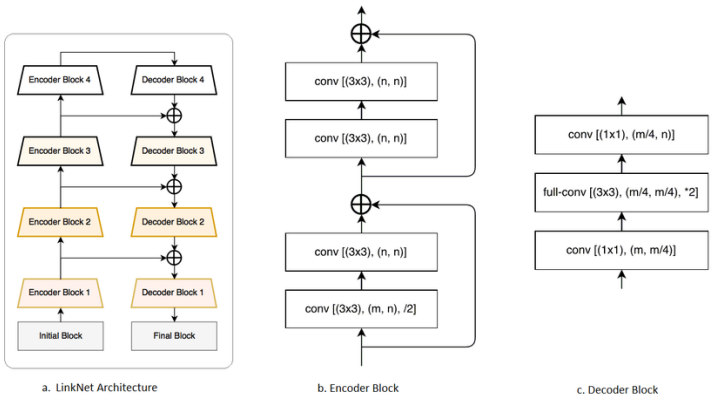


Figure 4: Example LinkNet architecture, image taken from [1]. $/*2$ refers to an upsampling stride of 2 whereas $//2$ refers to a downsampling stride of 2.

Block	Encoder		Decoder	
	m	n	m	n
1.	64	64	64	64
2.	64	128	128	64
3.	128	256	256	128
4.	256	512	512	256

Figure 5: The number of input channels (filters) m and output channels n used at each encoder and decoder block.

LinkNet uses ResNet as its encoder. ResNet consists of multiple encoder blocks where information is not only fed through convolutions but also added to the previous inputs using shortcut connections (see Figure 4). This is called *residual learning*. [5] LinkNet starts with an initial block that first performs convolution with kernel size 7 using stride 2 and 64 output channels followed by a (3×3) max-pooling with stride 2. The output block performs a deconvolution with 32 filters and stride 2, followed by a normal convolution with 32 filters and finishes with (2×2) deconvolution with stride 2 and a single output channel corresponding to the grey-scale image format. The number of input channels m and output channels n used at each encoder block is shown in Figure 5.

The networks share the property that the encoders send data to the decoders as to preserve some information that may be lost during downsampling. They are on the other hand dissimilar in the sense that in U-Net the encoder outputs are concatenated with the corresponding decoder inputs, thus increasing the number of data channels at each upsampling block, whereas in LinkNet encoder outputs are simply added element-wise to the decoder inputs using shortcut connections. For both models, the number of filters at each encoder block was increased by a factor two up until the bottleneck, after which the the number of filters were subsequently decreased by the same factor during the upsampling phase.

For both networks, images had to be loaded sequentially since an attempt to load the entire dataset would overload the working memory of the computer. An input pipeline was therefore used to load and preprocess the data in batches, utilising the CPU while simultaneously training the network on the GPU. The network was implemented on a system with an overclocked AMD Ryzen 5 2600 4.0GHz processor, 16GB of RAM and an NVIDIA GeForce RTX 2070 GPU.

DICE score was used to measure the network segmentation performance and is defined as:

$$DICE = \frac{2TP}{2TP + FP + FN},$$

where TP denotes the number of pixels that comprised true positives, i.e. pixels that were labelled as tumours and actually were tumours, whereas FP denotes false positives and FN false negatives.

5 Experiments

5.1 General setup

70% of the data was reserved for training, 20% for validation and finally 10% for testing. The learning rate was reduced by a factor of 10 when the validation loss hit a plateau. Both networks were trained using the Adam-optimiser and He-initialisation for 30 epochs each. Binary cross-entropy was used as loss function and the aforementioned DICE score was used to track the network accuracy. Images were also shuffled before every epoch.

5.2 U-Net

We experimented with the number of initial filters and settled for 16. U-net was found to perform well when using a batch size of 8. The initial learning rate was set to 10^{-4}. The learning curves are shown in Figure 6.

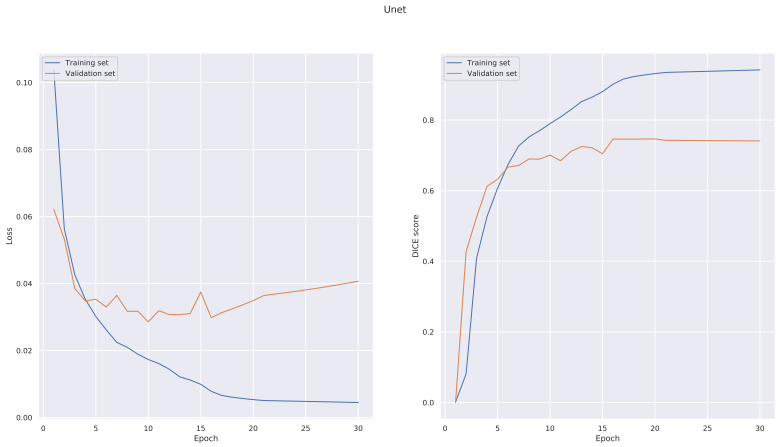


Figure 6: Loss function and DICE score as a function of epochs for U-Net.

The loss plot shows an indication that the data is overfitting, with the validation line increasing after epoch 12. This is also reflected in the DICE score plot.

5.3 LinkNet

We experimented with the number of start filters and found that 64 seemed to give the best results. We noticed that LinkNet seemed to be sensitive to parameter settings. Based on the parameter values used in [1], LinkNet was trained using an initial learning rate of $5 \cdot 10^{-4}$ and a batch size of 10. Batch normalisation was used after each convolutional layer. We also tried adding L_2-regularisation but that did not seem to significantly improve training. The learning curves are shown in Figure 7.

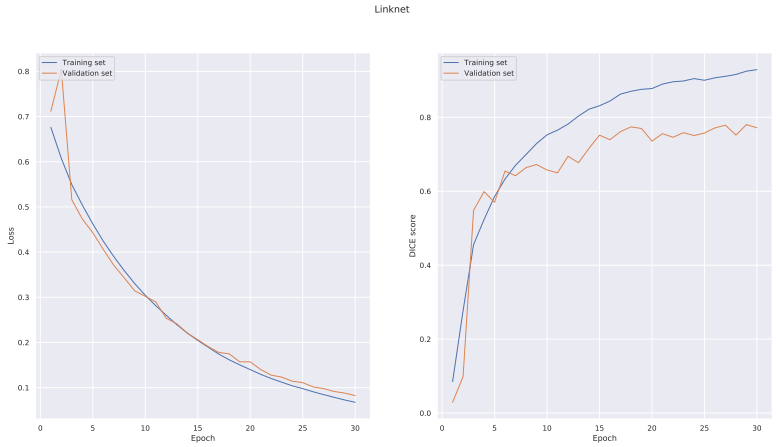


Figure 7: Loss function and DICE score as a function of epochs for LinkNet.

Comparing this plot to Figure 6 we see that U-Net tends to overfit the data slightly more than LinkNet.

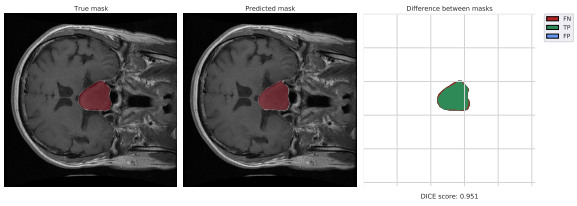
5.4 Final results

The final DICE scores achieved by both networks on the test set are shown in Table 1 and example predictions are shown in Figure 8.

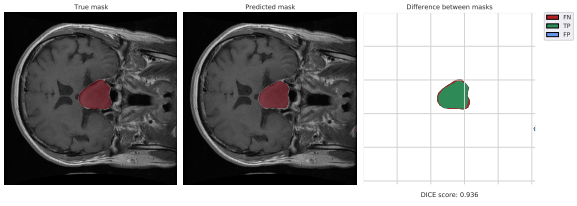
Method	**DICE**
U-Net	0.703
LinkNet	0.715

Table 1: Final DICE scores on test data.

Figures 8 and 9 show some of the results obtained from each network.



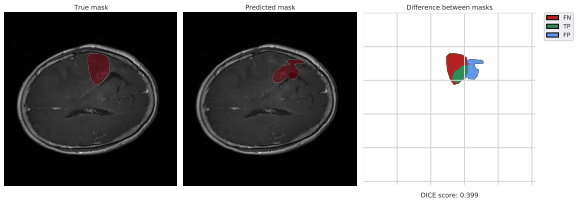
(a) Prediction made by U-Net.



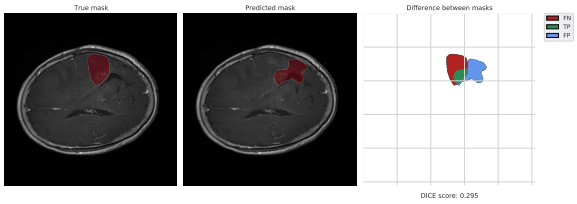
(b) Prediction made by LinkNet.

Figure 8

Both of the networks were able to make predictions with significant accuracy which is clear by the mask difference plot.



(a) Prediction made by U-Net.



(b) Prediction made by LinkNet.

Figure 9

There were however a few images where both networks were unable to make accurate predictions. A visual inspection of the image may give some indication as to why these predictions were not as successful. In the image in Figure 9 there is no obvious visual boundary between the brain tissue and the tumour for the network to pick up on. The part that is correctly identified, indicated by the green area, stand out against the rest of the image, whereas the rest of the tumour more similar to the healthy tissue is difficult to pick up on.

6 Conclusion

Our networks fall short by just a few decimal points of the target DICE score of 0.73. We consider this a success, given the scope and time constraints of the task.

Improvements could perhaps be made by implementing data augmentation into the preprocessing since there are relatively few samples (3064) in the used data set. The effective number of samples seen by the network during training may be increased by applying transformations to the available training data. Examples of such transformations would be changing the brightness, saturation or orientation of the image. This augmentation of the images introduces variations which the network has to adapt to and also works as regularisation.

Although this report has shown that it is relatively easy to detect brain tumours using CNNs, one must use these models with caution. When diagnosing and treating brain tumours, these results should be used as a tool and not as a replacement for medical professionals.

References

[1] Abhishek Chaurasia and Eugenio Culurciello. “LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation”. In: *2017 IEEE Visual Communications and Image Processing (VCIP)* (Dec. 2017), pp. 1–4. DOI: 10.1109/VCIP.2017.8305148. arXiv: 1707.03718. URL: http://arxiv.org/abs/1707.03718 (visited on 05/17/2020).

[2] Jun Cheng et al. “Enhanced Performance of Brain Tumor Classification via Tumor Region Augmentation and Partition”. In: *PLOS ONE* 10.10 (2015). Publisher: Public Library of Science, e0140381. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0140381. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0140381 (visited on 05/07/2020).

[3] N. Engl and J. Med. *Rain T*. 2001.

[4] Mohammad Havaei et al. “Brain Tumor Segmentation with Deep Neural Networks”. In: *Medical Image Analysis* 35 (Jan. 2017), pp. 18–31. ISSN: 13618415. DOI: 10.1016/j.media.2016.05.004. arXiv: 1505.03540. URL: http://arxiv.org/abs/1505.03540 (visited on 05/06/2020).

[5] Kaiming He et al. “Deep Residual Learning for Image Recognition”. In: *arXiv:1512.03385 [cs]* (Dec. 10, 2015). arXiv: 1512.03385. URL: http://arxiv.org/abs/1512.03385 (visited on 05/17/2020).

[6] Ali Işın, Cem Direkoğlu, and Melike Şah. “Review of MRI-based Brain Tumor Image Segmentation Using Deep Learning Methods”. In: *Procedia Computer Science*. 12th International Conference on Application of Fuzzy Systems and Soft Computing, ICAFS 2016, 29-30 August 2016, Vienna, Austria 102 (Jan. 1, 2016), pp. 317–324. ISSN: 1877-0509. DOI: 10.1016/j.procs.2016.09.407. URL: http://www.sciencedirect.com/science/article/pii/S187705091632587X (visited on 05/15/2020).

[7] B. H. Menze et al. “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)”. In: *IEEE Transactions on Medical Imaging* 34.10 (2015), pp. 1993–2024.

[8] Andriy Myronenko. “3D MRI brain tumor segmentation using autoencoder regularization”. In: *arXiv:1810.11654 [cs, q-bio]* (Nov. 19, 2018). arXiv: 1810.11654. URL: http://arxiv.org/abs/1810.11654 (visited on 05/12/2020).

[9] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *arXiv:1505.04597 [cs]* (May 18, 2015). arXiv: 1505.04597. URL: http://arxiv.org/abs/1505.04597 (visited on 05/11/2020).

[10] Zahra Sobhaninia et al. “Brain Tumor Segmentation Using Deep Learning by Type Specific Sorting of Images”. In: *arXiv:1809.07786 [cs, eess]* (Sept. 20, 2018). arXiv: 1809.07786. URL: http://arxiv.org/abs/1809.07786 (visited on 05/06/2020).

[11] Feifan Wang et al. “3D U-Net Based Brain Tumor Segmentation and Survival Days Prediction”. In: *arXiv:1909.12901 [cs, eess]* (Mar. 31, 2020). arXiv: 1909.12901. URL: http://arxiv.org/abs/1909.12901 (visited on 05/17/2020).