A Survey on Brain Image Segmentation Using Convolutional Neural Networks

Abstract

Automatic brain segmentation from MR images has become one of the major areas of medical research. Deep learning algorithms, specially convolutional neural networks (CNN), have been widely used for determining the exact location, orientation, and area of the lesion. This survey provides an overview of CNN-based approaches used for brain image segmentation. We describe and compare several state-of-the-art techniques for improving segmentation accuracy and performance.

1 Introduction

Image segmentation is one of the most important tasks in medical image analysis and is often the first and the most critical step in many clinical applications. In brain MRI analysis, image segmentation is commonly used for measuring and visualizing the brain's anatomical structures, for analyzing brain changes, for delineating pathological regions, and for surgical planning and image-guided interventions. In the last couple of years, automatic segmentation using deep learning have become the methodology of choice for analyzing medical images. These methods achieve the state-of-the-art results and proved to address medical problems better than other methods. In this survey, we review CNN, one of the most popular methods in deep learning, for brain MRI segmentation.

In this survey, first, we briefly review background knowledge on the MR imaging and previous works on brain segmentation. Section 3 focuses on two main neural network formations used in this field of research. In Section 4 we discuss different segmentation approaches using CNNs and briefly describe new extensions to traditional CNNs. Section 5 introduces state-of-the-art techniques for improving segmentation performance. Different evaluation methods used for analyzing segmentation results will be introduced in Section 6. Finally, in conclusion, we assess the impact of proposed methods and provide future directions for development.

2 Background

Recently, the rapid development of brain imaging technologies has helped in analyzing and studying the brain anatomy and function. Magnetic resonance imaging (MRI) has made an enormous progress in accessing brain injury and exploring brain anatomy. Usually, MRI images are presented in slices from top to bottom in 2 dimensions. However, nowadays using sophisticated computer calculation, these 2-dimensional slices can be joined together to produce a 3-dimensional model of the area of interest. Using 3D images instead of 2D images helps to precisely measure the volume of segmented tumor or lesion. In segmentation task, each **voxel** in a 3D image will be assigned to a class object.

Some of the most successful supervised brain segmentation methods are based on voxel-wise classifiers, such as random forests and Markov random field (MRF). Combining random forest with a generative Gaussian mixture model (GMM) will obtain tissue-specific probabilistic priors. In 2013, some researchers used MRF to incorporate spatial regularization.MRFs are commonly used to encourage spatial continuity of the segmentation. Although these methods have been very successful, it appears that their modeling capabilities still have significant limitations. At the same time, deep

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

learning techniques have emerged as a powerful alternative for supervised learning with great model capacity and the ability to learn highly discriminative features for the task at hand. In particular, Convolutional Neural Networks (CNNs) have been applied with promising results on a variety of biomedical imaging problems.

3 Network Formation

Modeling 3D volumetric image structures for segmentation is not a trivial task. Although computers' computational capabilities have improved significantly, still 3D CNNs for methods such as patch-wise -which will be discussed in next section- needs considerably high memory and computation. Moreover, the limited number of annotated 3D training images can cause some problems like overfitting or curse-of-dimensionality in CNNs. 2D CNNs are helpful to address these issues. However, in recent studies, 3D CNNs are used vastly according to the novel methods -will be discussed in next section- which overcome challenges of using 3D input images.

3.1 2D and Hybrid 2D/3D

2D image analysis by CNN is based on 2D CNNs, which is used extensively in computer vision applications on natural images. The question that arises, is how to process 3D MRI inputs as 2D inputs. Segmentation of a 3D brain scan is achieved by processing each 2D slice independently. Therefore, multiple works use 2D CNNs on three orthogonal 2D patches [12][15]. In [12] each 2D network learns to classify the same center voxel, viewed from an axial, sagittal and coronal perspective. Combining this ensemble of 2D networks enabled the segmentation method to become 3D aware. Some of the 2D methods will be briefly mentioned.

A 3D patch of size $a \times a \times a$ centred on the voxel is used to capture local information at a high level of detail. Then, three 2D orthogonal patches of size $b \times b$ (each extracted from the sagittal, coronal and transverse planes respectively), also centred on the voxel, are added with the purpose of capturing a slightly broader but still local context around the voxel of interest (VOI). The advantage of this method is that they capture 3D information with a significantly smaller amount of memory for storage than a dense 3D patch, allowing bigger patch sizes to be used [5].

In [20] for each point x of interest, they extract a multi-channel patch $P(x)$ around it, which has spatial dimensions $d1, d2, d3$. Here, $d1$ and $d2$ are taken to be in-slice dimensions corresponding to high resolution, and $d3$ is the lower-resolution axial direction. If we have N number of channels, instead of using $d1 \times d2 \times d3 \times N$, which is a 3D CNN we can have this $d1 \times d2 \times (N \times d3)$, which combines number of channels and the third dimension. Now we can have the simple 2D CNN. The justification for this step is that due to the lower resolution in $d3$ dimension, we expect that omitting the convolution in this direction will have a minor impact on accuracy. Also, [6] takes advantage of the same issue that the brain image datasets mostly lack resolution in the third dimension. They consider performing the segmentation slice by slice from the axial view.

Paper [18] proposed a 2.5D, referred as 2D/3D hybrid representation, which decomposes 3D volumes of interest into a set of random 2D orthogonal views via scale, random translation and rotations with respect to VOI centroid coordinates. This dimension reduction approach, increases data variation and prevents overfitting in CNN.

3.2 3D CNN

Since most of the medical images datasets were gathered in 3D format, it is challenging to apply 2D CNNs on 3D data because convolutions in a 2D CNN can only capture 2-dimensional spatial information, and neglect the information along the third dimension. To tackle this problem, the idea of 2D CNN has been extended to 3D CNN, which learns hierarchical spatial 3D features that can be used for classification or segmentation tasks. Figure 1 shows difference between 2D and 3D CNN [21]. Also, to address 3D CNN's computation cost, several methods have been developed, which will be discussed in next section.

A 3D CNN model is trained with a training set of 3D MR images to obtain 3D spatial probability maps, which indicates the likelihood of VOI voxels belonging to each class. In this model, the input has the form of a 3D image patch P and the CNN is formed by alternating 3D convolutional and

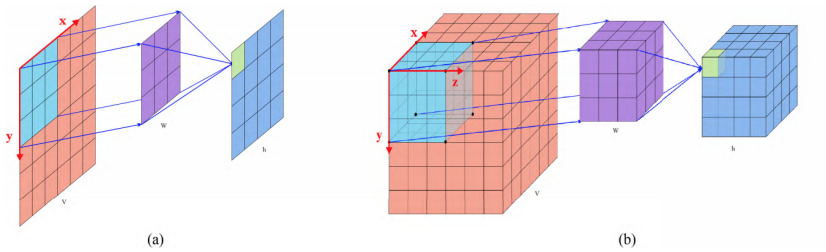

Figure 1: Difference between (a) 2D convolution and (b) 3D convolution [21].

pooling layers. These layers might be attached to fully connected layers [10] or a graphical model like CRF [8] or MRF [14], which finally generates probabilities that indicate the observed VOI voxel belongs to the class.

A Comparison between experimental results of 2D CNNs and 3D CNNs on BRATS and ISLES datasets indicates that the capabilities of 3D CNNs in capturing 3D patterns exceed those of 2D networks. Also, 3D CNNs are more efficient at inference time, which allows their adoption in a variety of research and clinical settings [8].

3.3 Number of CNN Layers

After primary papers as [20], which uses only one CNN layer, other works such as [8][6][12] used more than one layers of CNN and cascaded the CNN layers. In cascaded CNNs, we simply give the output of previous layers to the next layer. There are other methods proposed to feed the output probabilities of a first CNN as additional inputs to the layers of a second CNN. In this way, CNN architectures exploit the efficiency of CNNs and also models the dependencies between adjacent labels in the segmentation directly. In [6] the output of last CNN layers was used in three different ways. First, they concatenate the input of second layers with the output of the previous layer. Afterwards, they move up one layer in the local pathway and perform concatenation to its first hidden layer. Finally, they move to the very end of the second CNN and perform concatenation right before its output layer. In conclusion, according to the results, the first method gives the best result.

4 Approach

One of the most important issues in applying CNNs for segmentation, classification, registration etc, is how to manage the input data. Initially, patch-wise methods were used for this matter. By introducing Fully Convolutions Networks (FCN), a new era started in the field of image segmentation, which was continued with the appearance of methods such as U-Net architecture vastly known in the community of medical image analysis. Moreover, innovative methods like ResNet further improved the effectiveness of deep learning approaches by facilitating training of neural networks with higher depth.

4.1 Patch-wise

In the patch-wise method, CNN is applying in a sliding-window fashion in 3D or 2D space. At each point x, CNN is applied on a patch $p(x)$ around x. Given $p(x)$ to the CNN, it makes a prediction for the class of central patch point x. This method needs heavy computations and a huge memory space. The proposed method in [6] predicts the class of a pixel by processing the $M \times M$ patch centered on that pixel. The input of their CNN model is thus an $M \times M$ 2D patch with several modalities. They also use valid-mode convolution, meaning that the filter response (which filter size is $N \times N$) is not computed for pixel positions that are less than $N/2$ pixels away from the image border. Work of [15] used patch-wise method as patches are extracted in the axial slices and are normalized to have zero mean and unit variance in each sequence; the mean and variance are calculated in each sequence using all training patches. Figure 2 [1] indicates a patchwise model using $N \times N$ filter size.

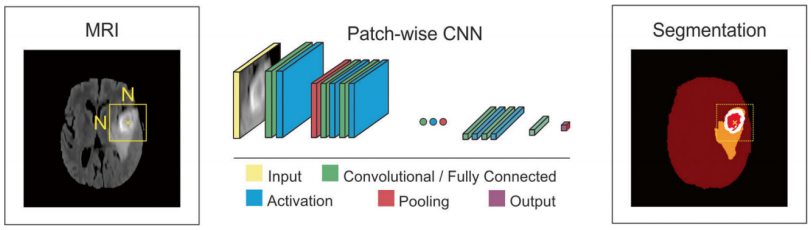

Figure 2: Schematic of a patch-wise CNN architecture for brain tumor segmentation task [1]

4.2 Fully Convolutions Neural Networks

In the common patch-wise classification setting, an input patch of size A is provided and the network outputs a single prediction for its central voxel. A drawback of this naive sliding-window approach is the huge overlap of neighboring pixels in the input patches and repeated computations of the convolutions. Fortunately, the convolution and dot product are both linear operators, and thus inner products can be written as convolutions and vice versa. By rewriting the fully connected layers as convolutions, the CNN can take input images larger than previous training images and produce a likelihood map, rather than an output for a single pixel. The resulting fully convolutional neural network (FCN) [11] can then be applied to an entire input image or volume in an efficient fashion.

The first paper [11] to propose FCN used different CNN layers, which made the picture 32 times smaller. After this downsampling step, they use upsampling to generate the picture in its original size. They had 16 and 8 versions that perform better than the initial 32 version. In this versions, they take the outputs in the middle of the CNN layers and concatenate the outputs in last CNN layers.

One of the problems of FCN is losing resolution in downsampling. Shift-and-stitch is one of the several methods proposed to prevent this decrease in resolution. The FCN is applied to shifted versions of the input image. By stitching the result together, one obtains a full resolution version of the final output, minus the pixels lost due to the valid convolutions.

In FCN deconvolution layers are mostly used as a upsampling method. Since upsampling with factor f is a convolution with a fractional input stride of $1/f$. As long as the f is integral, natural way to upsample is deconvolution with an output stride of f. Such an operation is trivial to implement since it simply reverses the forward and backward passes of convolution. The main advantage of the deconvolution filter is that it can be learned and doesn't need to be fixed. Figure 3 indicates a simple FCN model with convolution and deconvolution layers.

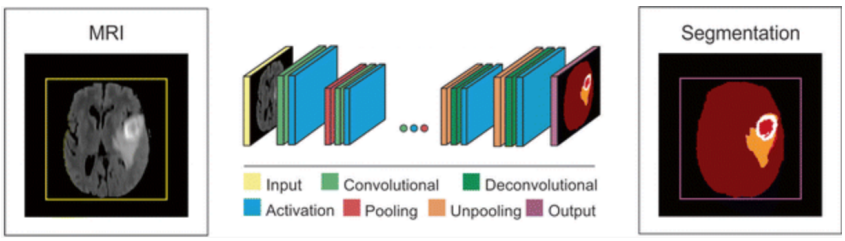

Figure 3: FCN model with deconvolution layers [1]

Recently, paper [8] proposed dense training as a hybrid model, which is neither patch-wise training nor full image input. Random patches are extracted from the training images. A batch is formed out of B of these samples. According to mathematics proposed in the paper, effective batch size will give us less computational and memory use.

4.3 Beyond CNNs

Despite the great power of 3D CNNs in object segmentation from volumetric data, these models may suffer from limited representation capability due to a shallow depth. Increasing depth of network

may cause gradient vanishing or optimization degradation. Recently, **Residual Deep Learning** has been widely used to tackle these problems by approximating the objective with residual functions.

In [3] a deep voxelwise residual network, referred as *VoxResNet*, was proposed, which consists of stacked residual modules referred as VoxRes module, with a total of 25 volumetric convolutional layers and 4 deconvolutional layers. This model tackles object segmentation tasks from high-dimensional volumetric images efficiently by learning an additive residual function with respect to the input feature. Figure 4 shows the architecture of the proposed model along with the VoxRes module.

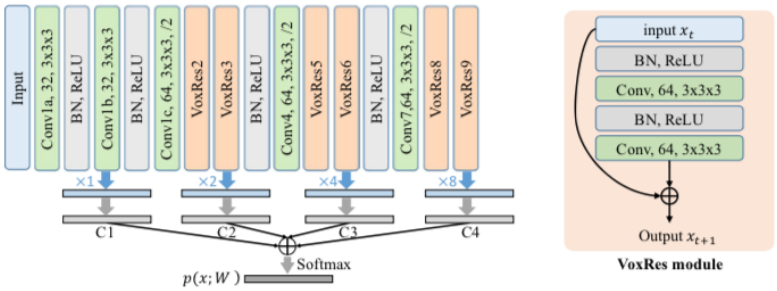

Figure 4: VoxResNet architecture for volumetric image segmentation [3]

Recently, an extension of ResNets has been introduced as ***DenseNet***, which are built from dense blocks and pooling operations, where each dense block is an iterative concatenation of previous feature maps. The layer connectivity pattern of these networks guarantees maximum information flow between layers in the network by connecting each layer to every other layer in the network. DenseNets can alleviate the vanishing-gradient problem, strengthen feature propagation, encourage feature reuse, and substantially reduce the number of parameters [7].

The other problem with 3D CNN models is their desperate need to the huge amount of data, which is not possible in medical image datasets. **U-Net** convolutoional neural network is a developed version of fully connected convolutional network, which is capable of working with very few training images and has high accuracy.

Paper [17] designed a U-Net convolutional architecture, which consists of a contracting path to analyze the whole image context and a successive expanding path to produce a full-resolution segmentation. Figure 5 illustrates proposed network architecture. This network gets 3D volumes as input and after passing through several 3D operations such as 3D convolutions, 3D max pooling, and 3D deconvolutional layers, the final layer maps feature vectors to desired classes. In this model, contracting path has the architecture of 3×3 convolutions followed by ReLU and max-pooling layers for downsampling. The expanding path consists of upsampling 2×2 convolution, which divides the number of feature channels by 2. Each channel is concatenated with corresponding cropped feature map from contracting path, and two 3×3 convolutions, each followed by a ReLU.

5 Improvements

Despite the significant influence of proposed methods on improving segmentation accuracy and precision, there are still some challenges with brain MR segmentation tasks like limited patch size or fake results. Following techniques are recent extensions and modifications that can be applied to the CNNs and FCNs models to address these challenges.

5.1 Multi-Stream Multi-Scale

One of the most important problems in MR brain segmentation is huge number of computations. Though, to have an accurate detection we need more context. In order to increase the context, the straightforward way is to feed larger patches which even make the computational cost and memory requirement worse. Therefore, different improvement methods such as multi-stream (multi-scale) are proposed to obtain accurate results without increasing the patch size.

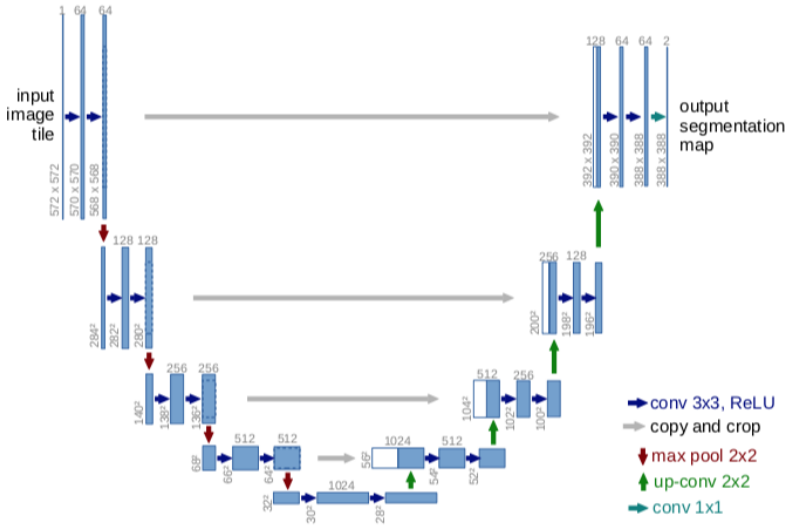

Figure 5: U-Net architecture proposed in [17]
. Left side is contracting path and right side is expansive path.

Multi-scaling can be described as resizing the input patch to different sizes and then passing them through CNN layers simultaneously (different path-ways) like [5], which used a downscaled patch spans by applying mean-pooling to the same region of the MRI as the original patch. Also, it can be defined as applying different kernel sizes on one patch in multiple different paths (path-ways) and at the end upsamplimg and stacking them together. This multi-scale approach allows the network to incorporate local details as well as global spatial consistency. Research of [8] used two-pathway, one of them with the actual size and the other one using downsampling and then upsampling. The result of the paths will be go through two layer fully connected neural network. Also [6] used two-pathway one with smaller 7×7 receptive fields and another with larger 13×13 receptive fields.

This method can go further and use more path-ways as [5] used 8 path-ways. In [13] different patch size with different kernel sizes in three different paths is used. This allows the weights and biases in the CNN to be specifically optimized for each patch size and corresponding kernel size. We decided to put the ensemble method of [12] also in this category as it gives the input to three different networks in parallel and ensembles the output which, is done by using the posterior probabilities of the last layer.

5.2 CRF and MRF

One challenge with segmentation approaches is that they sometimes lead to fake results. To resolve this, groups have tried to combine FCNs and CNNs with graphical models like MRFs and conditional random fields (CRFs) to refine the segmentation output. In most of the cases, graphical models are applied on top of the likelihood map, produced by CNNs or FCNs and act as label regularizers.

In 2016, research of [4] presented a cascaded 2D fully convolutional architecture along with a dense 3D conditional random fields (CRFs) for liver and lesion segmentation on CT dataset. Moreover, for MR brain segmentation [3] employed CRF as a post-processing step to refine the network's output. This CRF is capable of modeling arbitrarily large voxel-neighborhoods, but it is also computationally efficient and makes it ideal for processing 3D multi-modal medical scans.

In [19], they improved brain segmentation results by adding a Markov random field (MRF) on top of 2D CNN architecture. They considered the CNN output, which indicates the probability of each pixel belonging to a class label, as a potential for MRF and then they used an alpha-extension technique to minimize the energy corresponded to each possible label assignment. There are other categories of refinement methods such as [4], which the segmentation is refined using a cellular automaton-based seed growing method known as growcut. Describing details of these methods is beyond this survey's purpose.

6 Implementation & Evaluation

In brain image datasets, the classes are highly imbalanced, since there are much more samples of normal tissue than lesion tissue. To cope with this, different papers have proposed different methods. In [15], they extract around 40% of training samples from normal tissue, while the remaining 60% corresponds to brain tumor samples with approximately balanced numbers of samples across classes.

The other approach to overcome class imbalance is to change the objective function in a way that tumor voxels get higher weights than non-diseased classes. Authors of [2] defined a weighted sum of the mean squared difference of the tumor voxels and non-tumor voxels with a larger weight for the specificity to make it less sensitive to the data imbalance. As an alternative solution, [16] performed data augmentation on positive samples. They used all samples from the underrepresented classes and randomly sampled from the others to balance the dataset.

For evaluation, different volumetric measures are used to compare a predicted brain segmentation with the ground truth segmentation mask. Among, different measures, the *Dice score* is the most widespread measure used for the comparison of two segmentation and it has the form:

$$D = \frac{2|P \cap R|}{|P| + |R|} \tag{1}$$

where P is predicted brain segmentation and R is ground truth segmentation.

Positive Predictive Value (PPV), Sensitivity and Robust Hausdorrf distance are other popular metrics for segmentation evaluation. Also, computing the False Positive, False Negative and absolute error maps between the ground truth and predicted mask are other common ways to analyze segmentation results [9].

7 Discussion and Conclusion

From more than 20 papers that we reviewed in this survey, it is obvious that deep learning-based models outperform other models in medical image analysis. In early works, due to limited memory and computation capabilities of computers, 3D images were transformed to 2D space and were fed into 2D CNNs in patchwise manner. With the advent of FCNs and development of GPUs, memory limitation was no longer a problem in medical image analysis. FCNs with 3D convolutions, which are capable of processing a whole 3D image at a time, have been widely used for segmentation and classification tasks. Migrating from 2D CNNs to 3D CNNs greatly influenced networks' information representation capabilities and consequently their accuracy. Also, using ResNet and U-Net architectures alleviates gradient vanishing and small datasets issues in such models.

The other challenge with 3D segmentation is the limited number of samples for training deep networks, which sometimes lead to overfitting or unreliable responses. To combat this, groups have tried to apply graphical models like MRFs or CRF on top of the CNNs or FCNs. Class-imbalance is the other data-related challenge in medical imaging. The majority of training samples are normal and do not contain any special information. To address this problem, different strategies like data augmentation, weighted loss function and sampling positive class have been adopted.

Summarizing, CNN models and their extensions have obtained promising results in medical image segmentation tasks. Most of current models are based on supervised learning, which requires a huge amount of effort on labeling training examples. Using unsupervised methods is more data efficient since they can be trained with the wealth of unlabeled data available in the world. So far, current methods were randomly initialized and trained on a limited data. Transfer learning could be used to share deep learning models, which are perfectly trained on big datasets and improve the generalization ability of these models across datasets with less effort than learning from scratch.

References

[1] Zeynettin Akkus, Alfiia Galimzianova, Assaf Hoogi, Daniel L Rubin, and Bradley J Erickson. Deep learning for brain mri segmentation: state of the art and future directions. *Journal of digital imaging*, 30(4):449–459, 2017.

[2] Tom Brosch, Youngjin Yoo, Lisa YW Tang, David KB Li, Anthony Traboulsee, and Roger Tam. Deep convolutional encoder networks for multiple sclerosis lesion segmentation. In

International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 3–11. Springer, 2015.

[3] Hao Chen, Qi Dou, Lequan Yu, and Pheng-Ann Heng. Voxresnet: Deep voxelwise residual networks for volumetric brain segmentation. *arXiv preprint arXiv:1608.05895*, 2016.

[4] Patrick Ferdinand Christ, Mohamed Ezzeldin A Elshaer, Florian Ettlinger, Sunil Tatavarty, Marc Bickel, Patrick Bilic, Markus Rempfler, Marco Armbruster, Felix Hofmann, Melvin D'Anastasi, et al. Automatic liver and lesion segmentation in ct using cascaded fully convolutional neural networks and 3d conditional random fields. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 415–423. Springer, 2016.

[5] Alexandre de Brébisson and Giovanni Montana. Deep neural networks for anatomical brain segmentation. *arXiv preprint arXiv:1502.02445*, 2015.

[6] Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, and Hugo Larochelle. Brain tumor segmentation with deep neural networks. *Medical image analysis*, 35:18–31, 2017.

[7] Gao Huang, Zhuang Liu, Kilian Q Weinberger, and Laurens van der Maaten. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, volume 1, page 3, 2017.

[8] Konstantinos Kamnitsas, Christian Ledig, Virginia FJ Newcombe, Joanna P Simpson, Andrew D Kane, David K Menon, Daniel Rueckert, and Ben Glocker. Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical image analysis*, 36:61–78, 2017.

[9] Jens Kleesiek, Gregor Urban, Alexander Hubert, Daniel Schwarz, Klaus Maier-Hein, Martin Bendszus, and Armin Biller. Deep mri brain extraction: a 3d convolutional neural network for skull stripping. *NeuroImage*, 129:460–469, 2016.

[10] Robert Korez, Boštjan Likar, Franjo Pernuš, and Tomaž Vrtovec. Model-based segmentation of vertebral bodies from mr images with 3d cnns. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 433–441. Springer, 2016.

[11] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[12] Mark Lyksborg, Oula Puonti, Mikael Agn, and Rasmus Larsen. An ensemble of 2d convolutional neural networks for tumor segmentation. In *Scandinavian Conference on Image Analysis*, pages 201–211. Springer, 2015.

[13] Pim Moeskops, Max A Viergever, Adriënne M Mendrik, Linda S de Vries, Manon JNL Benders, and Ivana Išgum. Automatic segmentation of mr brain images with a convolutional neural network. *IEEE transactions on medical imaging*, 35(5):1252–1261, 2016.

[14] Pim Moeskops, Jelmer M Wolterink, Bas HM van der Velden, Kenneth GA Gilhuijs, Tim Leiner, Max A Viergever, and Ivana Išgum. Deep learning for multi-task medical image segmentation in multiple modalities. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 478–486. Springer, 2016.

[15] Sérgio Pereira, Adriano Pinto, Victor Alves, and Carlos A Silva. Deep convolutional neural networks for the segmentation of gliomas in multi-sequence mri. In *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 131–143. Springer, 2015.

[16] Sérgio Pereira, Adriano Pinto, Victor Alves, and Carlos A Silva. Brain tumor segmentation using convolutional neural networks in mri images. *IEEE transactions on medical imaging*, 35(5):1240–1251, 2016.

[17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[18] Holger R Roth, Le Lu, Ari Seff, Kevin M Cherry, Joanne Hoffman, Shijun Wang, Jiamin Liu, Evrim Turkbey, and Ronald M Summers. A new 2.5 d representation for lymph node detection using random sets of deep convolutional neural network observations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 520–527. Springer, 2014.

[19] Mahsa Shakeri, Stavros Tsogkas, Enzo Ferrante, Sarah Lippe, Samuel Kadoury, Nikos Paragios, and Iasonas Kokkinos. Sub-cortical brain structure segmentation using f-cnn's. In *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*, pages 269–272. IEEE, 2016.

[20] Darko Zikic, Yani Ioannou, Matthew Brown, and Antonio Criminisi. Segmentation of brain tumor tissues with convolutional neural networks. *Proceedings MICCAI-BRATS*, pages 36–39, 2014.

[21] Liang Zou, Jiannan Zheng, Chunyan Miao, Martin J Mckeown, and Z Jane Wang. 3d cnn based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural mri. *IEEE Access*, 5:23626–23636, 2017.