3D Dense-U-Net for MRI Brain Tissue Segmentation

Martin Kolařík, Radim Burget, Václav Uher
Department of Telecommunications
BRNO UNIVERSITY OF TECHNOLOGY
Brno, Czech Republic

Malay Kishore Dutta
Dept. of Electronics & Communication Engineering
Amity University
Noida, India

***Abstract*—This paper presents a fully automatic method for 3D segmentation of brain tissue on MRI scans using modern deep learning approach and proposes 3D Dense-U-Net neural network architecture using densely connected layers. In contrast with many previous methods, our approach is capable of precise segmentation without any preprocessing of the input image and achieved accuracy 99.70 percent on testing data which outperformed human expert results. The architecture proposed in this paper can also be easily applied to any project already using U-net network as a segmentation algorithm to enhance its results. Implementation was done in Keras on Tensorflow backend and complete source-code was released online.**

***Keywords*—3D segmentation; brain; deep learning; imageprocessing; mri; neural networks; opensource; semantic segmentation; u-net**

I. Introduction

Magnetic resonance imaging is modern and widely used radiology method to display the internal organs of the human body. Using MRI, it is possible to obtain slices of a certain area of the body, which can be further processed and linked up to the resulting 3D image of the desired organ. The advantage of MRI over the other imaging methods in diagnostic radiology is greater precision in the majority of organs, which is a result of different signal strengths in different soft tissues. This characteristic makes it an ideal diagnostic tool for imaging human brain - either its tissue for volumetry or for finding and measuring abnormal regions.

Segmentation of human brain tissue on MRI scan is a time consuming task even for an experienced radiologist. Therefore the automation of this process is very valuable and can help doctors determine the correct diagnose faster when they are presented with precisely segmented brain within few seconds. Problem with automatic segmentation of any tissue in medicine is that the solution must be reliable and at least as precise as when doctor would have done the same task on his own. Our solution provides good results on real data and is a small step towards helping the doctors automate the routine and lengthy tasks and to provide objective features about patient progress.

Research described in this paper was financed by the Ministry of the Interior of the Czech Republic by the grant VI20172019086 and the National Sustainability Program under grant LO1401. For the research, infrastructure of the SIX Center was used.

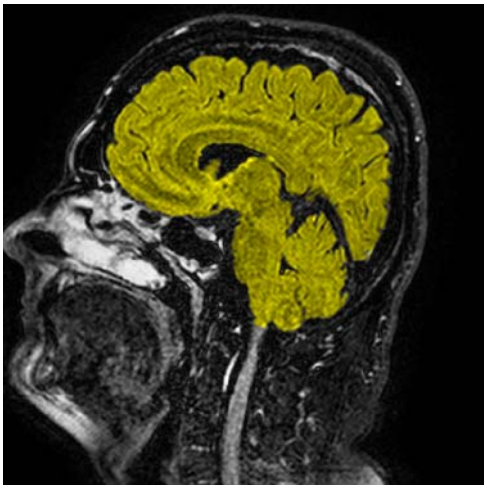

Fig. 1. Example of MRI sagitall brain scan slice - brain tissue segmented with our system is highlighted in yellow.

The rest of the paper is structured as follows. In the section II we analyze existing methods of brain tissue segmentation and also briefly list related works in the field of deep learning segmentation in general. In section III we described all information about data used for training and evaluating results of our proposed networks. In subsection B of section III we described how we prepared the data as input of neural network and how were we able to deal with limited amount of GPU memory. Section IV contains all the information about neural networks architectures and we included link to our github repository with source code, so this work can be easily repeated. In the section V we interpret our results and describe used metrics.

II. Related works

A. Brain segmentation methods

The problem of automatic brain tissue segmentation has been very well explored before deep learning algorithms became a standard for semantic segmentation. Despotović et.al [1] covers an overview of older methods such as thresholding, clustering or some form of simpler machine learning algorithms. These methods also rely heavily on image preprocessing as in [2] therefore they are not automatic. One of quite

frequently used preprocessing method is some form of skull-stripping algorithm, which removes the bone tissue of skull from the input picture before processing as in [3]. Since deep learning methods achieve higher accuracy even without any preprocessing of input image, most of the ongoing research including this paper is now using some form of deep neural network for brain segmentation. Overview of the current state of the art in the field of brain segmentation using deep learning can be found in [4].

B. Deep learning segmentation methods

As we stated in previous subsection, current state of the art methods for segmenting image data are deep learning based. Comparison of currently used network architectures is in [5]. Shelhamer et.al. [6] proposed first fully convolutional architecture for segmentation. One of the most popular architectures is autoencoder type of network called U-Net [7]. It consists of downsampling, bridge and upsamling blocks and is widely used for segmenting medical data because its ability to be properly trained using little data comparing to its competitors such as [6]. Development in the area of classification networks using data interconnections from coarse to fine layers of the network as in residual networks [8] and followed by densely connected networks [9] led to application of this principle also in segmentation networks. Popular segmentation architecture using these principles is Tiramisu network [10]. All cited architectures are designed for 2D data.

Our goal was to design U-net architecture with densely connected layers for 3D data processing and compare its accuracy with results of the original U-Net implementation and U-Net with added residual connections.

III. Dataset

A. Data overview

Dataset consisted of 22 brain scans from different patients and each scan contained 257 sagitall slices of human brain. MRI data were provided by the Department of Radiology from The University Hospital Brno and all original slices are labeled by two independent human experts resulting into two sets of ground truth masks suitable for semantic segmentation labeling all pixels either as brain tissue (white) or non-brain tissue (black).

Fig. 2. Reference ground truth mask labeled by human expert (left), mask labeled by our system (right).

The first expert labeled the data very precisely and accurately and these labels are used as a reference data for training and evaluation. The second expert labeled the data as it is done regularly on everyday basis in medical praxis. Example of reference ground truth mask and mask labeled by our system can be seen in Fig. 2.

B. Data preparation

We used 21 scans as a training data, 10 percent of that data served for validation. Last remaining 22nd brain scan was used for testing data. Due to the fact that we used 3D segmentation and we wanted to use as much slices as possible at one batch to keep the spatial information but still be able to train complex neural network architectures on currently available resources (GPU gtx 1080ti with 11 GB of RAM), we had to divide the data to batches of 16 slices each. Due to the fact that approximately 15 slices on both ends of every subject consisted of only non brain tissue and therefore contained very little information value, we discarded last slice of every subject and used only 256 slices from each scan so the final number is divisible by 16. In order to fully utilize the benefits of 3D segmentation we divided the training dataset to overlapping batches. Example can be seen on Fig. 3

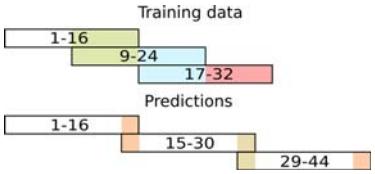

Fig. 3. Example of training and prediction data batch overlapping. Numbers show which slices does each batch contain.

We wanted to use the same overlapping data technique for outcome predictions from the system. But to lower the time the systems needs to predict the results, we overlapped only 2 pixels on each end of every batch. Prediction then consists of central 14 slices of each batch. This results in 252 testing images used for prediction.

IV. Methodology

A. Neural network architectures

In this paper we introduced 3D Dense-U-Net network which is based on original U-Net implementation [7] but with added interconnections between layers processing the same feature size, its model is in Fig. 4. This helps the network to achieve faster learning curve and obtain higher level of details. It is an autoencoder type architecture [7] with 4 down-sampling and 4 up-sampling blocks which are connected by a bridge block (in Fig. 4. the most lower part of the network). Feature size is halved after passing through every down-sampling block by a maxpooling layer at its bottom. On the beginning of every up-sampling block the feature size is doubled using transposed convolution layer as desribed by [11] with stride of size 2 for each dimension. Passing information through interconnections is possible by using zero padding. After every convolution

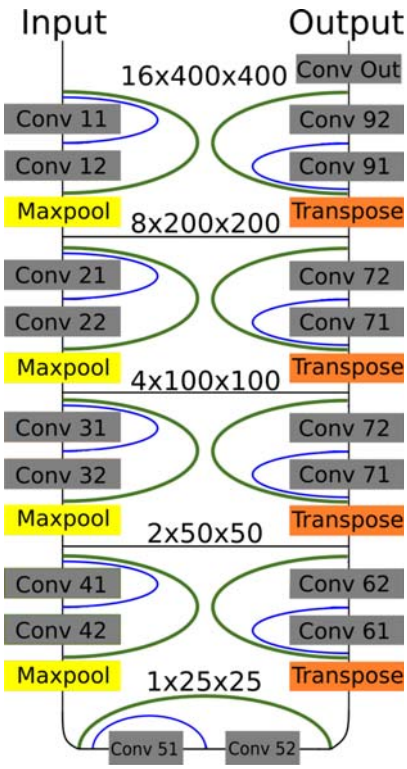

Fig. 4. Dense-U-Net network model. Residual interconnections are in green color, dense interconnections in blue.

TABLE I. NUMBER OF NEURONS IN EACH LAYERS.

Input		Output	
Layers	Neurons	Layers	Neurons
11,12	32	91,92	32
21,22	64	81,82	64
31,32	128	71,72	128
41,42	256	61,62	256
Bridge block – Convolutions 51,52 – 512 neurons each			

operation we fill the newly computed vector with zeros to its original length and therefore both convolution layers use the same size of the input vector.

In our experiment we compared results of original U-Net and Residual-U-Net with added only residual interconnection (green in Fig. 4). Especially Residual-U-Net network had very good results taking into account its number of parameters.

B. Implementation details

Complete source code used in this paper is available at[1]. All neural networks architectures were trained using Keras framework [12] on Tensorflow backend. Training was done on nvidia gtx 1080ti graphics card with 11 GB of memory using CUDA 8.0. As an inspiration for first U-Net model was used repository [13]. Our basic U-Net architecture uses 3D data processing layers. All networks use a binary cross-entropy as a loss function. We used Adam optimizer with parameters learning_rate=$1e-5$, beta1=0.9, beta2=0.999, epsilon=$1e-08$

[1]https://github.com/mrkolarik/3D-brain-segmentation

TABLE II. PARAMETERS OF NETWORK LAYERS - THEIR CONVOLUTION KERNEL SIZE, STRIDES AND ACTIVATION FUNCTION.

Downsampling block			
Layer type	Kernel (pool) size	Strides	Activation
Convolution 3d	3,3,3	0,0,0	Relu
Convolution 3d	3,3,3	0,0,0	Relu
Maxpooling	2.2.2	0,0,0	-

Upsampling block			
Layer type	Kernel (pool) size	Strides	Activation
Transpose Conv.	2.2.2	2,2,2	-
Convolution 3d	3,3,3	0,0,0	Relu
Convolution 3d	3,3,3	0,0,0	Relu

Bridge block			
Layer type	Kernel (pool) size	Strides	Activation
Convolution 3d	3,3,3	0,0,0	Relu
Convolution 3d	3,3,3	0,0,0	Relu

Output convolution			
Layer type	Kernel (pool) size	Strides	Activation
Convolution 3d	1,1,1	0,0,0	Sigmoid

and decay=$1.99e-7$. Using decay parameter we lower the learning rate parameter each epoch by constant value.

We created Residual-U-Net and Dense-U-Net architectures by adding interconnections to default 3D U-Net architecture. Residual-U-Net uses only interconnection over whole down or up-sampling blocks as can be seen in Fig. 4 and residual interconnection is green. Dense-U-Net uses both residual and dense interconnections (blue in Fig. 4) which passes unprocessed information to the middle layer of down and up-sampling blocks.

Network uses as input data value in range [0,1]. Data had to be converted on input and output accordingly, because original images were encoded as 8bit monochromatic images in png format and their pixel values were in range [0,255]. As we used sigmoid activation on output network layer, the network did not label all pixels only as 0 and 1, but also as some values in between. Therefore we used as only post-processing on predicted data thresholding. All pixels with value lesser than 0.5 were labeled as 0 and with value greater as 1.

We first trained all three networks for 50 epochs for evaluation purposes. Then we trained the Dense-U-Net for 99 epochs as for final results, because it achieved best results for all three evaluated networks.

V. RESULTS

We used five different metrics for evaluation so that we can easily compare our results to similar works. All metrics measured are compared to masks labeled by human expert as precise as possible. First used metrics is pixel accuracy, its equation is (1) where N_{TP} stands for true positive, N_{TN} true negative, N_{FP} false positive and N_{FN} false negative.

$$A_P(X,Y) = \frac{N_{\text{TP}} + N_{\text{TN}}}{N_{\text{TP}} + N_{\text{TN}} + N_{\text{FP}} + N_{\text{FN}}} \quad (1)$$

$$D(X,Y) = \frac{2 * |X \cap Y|}{|X| + |Y|} \quad (2)$$

TABLE III. COMPARISON OF TESTED U-NET VERSIONS. USED METRICS - PIXEL ACCURACY, DICE COEFFICIENT, INTERSECTION OVER UNION, AVERAGE HAUSDORFF DISTANCE [VOXEL] AND AREA UNDER ROC CURVE.

Metric	Dense-Unet	Res-Unet	Unet	Human
P.A.	0.99703	0.99662	0.99619	0.99489
Dice c.	0.98843	0.98686	0.98514	0.98033
I.o.U.	0.97713	0.97407	0.97072	0.96141
A.H.D.	0.01334	0.01911	0.02427	0.02479
A.u.R.C.	0.99439	0.99353	0.99205	0.98325

$$I_{oU}(X,Y) = \frac{|X \cap Y|}{|X \cup Y|} = \frac{|X \cap Y|}{|X| + |Y| - |X \cap Y|} \quad (3)$$

Dice coefficient [14] is in equation (2) and intersection over union [15], also known as Jaccard index is on equation (3). X and Y stand for a set of positive pixels on first and second compared mask. We used Visceral segmentation tool [15] for computing the results in all metrics.

As stated in section III., we computed the predictions on the last 22nd brain scan over all 252 slices. This scan was not used for training and therefore our results show reliability of our method on unseen data. We compared the predicted data to mask labeled by a human expert as in standard medical praxis, which are in table III. in column labeled "Human". Comparison of results of all three networks can be seen on the table III..

Dense-U-Net network had the best results in all metrics, but even segmenting using basic U-Net network gave better results than human expert labeled masks. We trained it again for 99 epochs to get final results which are in table IV..

Properly trained Dense-U-Net achieved over 0.2 percent better pixel accuracy than a human.

Results show us that the method is capable of achieving higher accuracy than human expert even on real data that were not part of training dataset.

VI. CONCLUSION

We introduced upgraded U-Net architecture with densely connected layers modified for 3D data processing, the 3D Dense-U-Net. This network achieved pixel accuracy on testing data 99.70 percent which exceeded human expert performance done as in standard medical praxis 99.49 percent accuracy.

Our approach can easily be applied to any segmentation method already using U-Net architecture. Resulting source-code was released and its link is in section IV..

Using data preparation technique described in chapter dataset, we were able to train a deep neural network with 3D data using GPU with 11 GB RAM.

TABLE IV. FINAL RESULTS OF DENSE-U-NET NETWORK COMPARED TO HUMAN EXPERT RESULTS. USED METRICS - PIXEL ACCURACY, DICE COEFFICIENT, INTERSECTION OVER UNION, AVERAGE HAUSDORFF DISTANCE [VOXEL] AND AREA UNDER ROC CURVE.

Metric	Dense-Unet	Human
P.A	0.99705	0.99489
Dice c.	0.98852	0.98033
I.o.U.	0.97731	0.96141
A.H.D.	0.01287	0.02479
A.u.R.C.	0.99467	0.98325

In future we plan to further upgrade data preparation technique so we will be able to train more densely interconnected architectures, plan to evaluate our method on other publicly available datasets and to design a more universal data preparation technique, which will be suitable even for number of input images not divisible by 16.

REFERENCES

[1] I. Despotović, B. Goossens, and W. Philips, "MRI Segmentation of the Human Brain: Challenges, Methods, and Applications"Computational and Mathematical Methods in Medicine Volume 2015 (2015), Article ID 450341, 23 pages

[2] V. Uher and R. Burget, "Automatic 3D segmentation of human brain images using data-mining techniques"2012 35th International Conference on Telecommunications and Signal Processing (TSP), pp. 578-580

[3] V. Uher, R. Burget, J. Masek, and M. K. Dutta, "3D brain tissue selection and segmentation from MRI"2013 36th International Conference on Telecommunications and Signal Processing (TSP), Rome, 2013, pp. 839-842., doi: 10.1109/TSP.2013.6614057

[4] Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin, and B. J. Erickson, "Deep Learning for Brain MRI Segmentation: State of the Art and Future Directions"Journal of digital imaging. 30., doi: 10.1007/s10278-017-9983-4.

[5] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A Review on Deep Learning Techniques Applied to Semantic Segmentation ", CoRR, vol. 2017, no. abs/1704.06857.

[6] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation "CoRR, vol. 2014, no. abs/1411.4038.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation "

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition "CoRR, vol. 2015, no. abs/1512.03385.

[9] G. Huang, Z. Liu, and K. Q. Weinberger, "Densely Connected Convolutional Networks "CoRR, vol. 2016.

[10] S. Jgou, M. Drozdzal, D. Vzquez, A. Romero, and Y. Bengio, "The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation ", CoRR, vol. 2016.

[11] NOH, Hyeonwoo, Seunghoon HONG and Bohyung HAN, "Learning Deconvolution Network for Semantic Segmentation "

[12] F. Chollet, "Keras ", GitHub, 2015, [Online]. Available: https://github.com/keras-team/keras

[13] M. Jocić, "Deep Learning Tutorial for Kaggle Ultrasound Nerve Segmentation competition, using Keras ", Github, 2017, [Online]. Available: https://github.com/jocicmarko/ultrasound-nerve-segmentation

[14] K. H. Zou, S. K. Warfield, A. Bharatha, C. M. C. Tempany, M. R. Kaus, S. J. Haker, W. M. Wells, F. A. Jolesz, and R. Kikinis, "Statistical validation of image segmentation quality based on a spatial overlap index1 "

[15] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool ", [Online]. Available: http://bmcmedimaging.biomedcentral.com/articles/10.1186/s12880-015-0068-x