

BRNO UNIVERSITY OF TECHNOLOGY

Faculty of Electrical Engineering
and Communication

MASTER'S THESIS

Brno, 2020

Usevalad Ustsinau



BRNO UNIVERSITY OF TECHNOLOGY

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

FAKULTA ELEKTROTECHNIKY
A KOMUNIKAČNÍCH TECHNOLOGIÍ

DEPARTMENT OF BIOMEDICAL ENGINEERING

ÚSTAV BIOMEDICÍNSKÉHO INŽENÝRSTVÍ

SEGMENTATION OF BRAIN TUMOURS IN MRI IMAGES USING DEEP LEARNING

SEGMENTACE NÁDORŮ MOZKU V MRI DATECH S VYUŽITÍM HLOUBKOVÉHO UČENÍ

MASTER'S THESIS

DIPLOMOVÁ PRÁCE

AUTHOR

AUTOR PRÁCE

Usevalad Ustsinau

SUPERVISOR

VEDOUCÍ PRÁCE

Ing. Jiří Chmelík

BRNO 2020

Master's Thesis

Master's study programme **Electrical, Electronic, Communication and Control Technology**
field Biomedical and Ecological Engineering
Department of Biomedical Engineering

Student: Usevalad Ustsinau

ID: 214309

Year of study: 2

Academic year: 2019/20

TITLE OF THESIS:

Segmentation of brain tumours in MRI images using deep learning

INSTRUCTION:

1) Do a literature review focused on deep learning methods for image segmentation, study and describe MRI techniques for brain imaging. 2) Propose a deep learning method for brain tumour segmentation in MRI data. 3) Implement the proposed method in the selected programming language. 4) Test the implemented algorithm on the available image database and optimize the method in terms of maximal segmentation effectiveness. 5) Compare achieved results with other authors. 6) Evaluate and discuss achieved results.

REFERENCE:

[1] GARCIA-GARCIA, A., S. ORTS-ESCOLANO, S. OPREA, V. VILLENA-MARTINEZ, P. MARTINEZ-GONZALEZ a J. GARCIA-RODRIGUEZ. A survey on deep learning techniques for image and video semantic segmentation. Applied Soft Computing. 2018, 70, 41-65. DOI: 10.1016/j.asoc.2018.05.018. ISSN 1568-4946.

[2] LITJENS, G., T. KOOI, B. E. BEJNORDI, et al. A survey on deep learning in medical image analysis. Medical Image Analysis. 42, 60-88. DOI: 10.1016/j.media.2017.07.005. ISSN 1361-8415.

Assignment deadline: 3. 2. 2020

Submission deadline: 29. 5. 2020

Head of thesis: Ing. Jiří Chmelík

prof. Ing. Ivo Provazník, Ph.D.
Chair of study branch board

WARNING:

The author of this Master's Thesis claims that by creating this thesis he/she did not infringe the rights of third persons and the personal and/or property rights of third persons were not subjected to derogatory treatment. The author is fully aware of the legal consequences of an infringement of provisions as per Section 11 and following of Act No 121/2000 Coll. on copyright and rights related to copyright and on amendments to some other laws (the Copyright Act) in the wording of subsequent directives including the possible criminal consequences as resulting from provisions of Part 2, Chapter VI, Article 4 of Criminal Code 40/2009 Coll.

ABSTRACT

The following master's thesis paper equipped with a short description of CT scans and MR images and the main differences between them, explanation of the structure of convolutional neural networks and how they implemented into biomedical image analysis, besides it was taken a popular modification of U-Net and tested on two loss-functions. As far as segmentation quality plays a highly important role for doctors, in experiment part it was paid significant attention to training quality and prediction results of the model. The experiment has shown the effectiveness of the provided algorithm and performed 100 training cases with the following analysis through the similarity. The proposed outcome gives us certain ideas for future improving the quality of image segmentation via deep learning techniques.

KEYWORDS

Medical Imaging, Brain Tumour, Convolutional Neural Network, Segmentation, U-Net

USTSINAU, Usevalad. *Segmentation of brain tumours in MRI images using deep learning*. Brno, 2020, 53 p. Master's Thesis. Brno University of Technology, Faculty of Electrical Engineering and Communication, Department of Biomedical Engineering. Advised by Ing. Jiří Chmelík

DECLARATION

I declare that I have written the Master's Thesis titled "Segmentation of brain tumours in MRI images using deep learning" independently, under the guidance of the advisor and using exclusively the technical references and other sources of information cited in the thesis and listed in the comprehensive bibliography at the end of the thesis.

As the author I furthermore declare that, with respect to the creation of this Master's Thesis, I have not infringed any copyright or violated anyone's personal and/or ownership rights. In this context, I am fully aware of the consequences of breaking Regulation § 11 of the Copyright Act No. 121/2000 Coll. of the Czech Republic, as amended, and of any breach of rights related to intellectual property or introduced within amendments to relevant Acts such as the Intellectual Property Act or the Criminal Code, Act No. 40/2009 Coll., Section 2, Head VI, Part 4.

Brno

.....

author's signature

ACKNOWLEDGEMENT

I would like to thank the advisor of my thesis Ing. Jiří Chmelík for his valuable comments, guidance and ideas for work. Besides this work was carried out with scientific support from Dr. Michael Goetz from the German Cancer Research Center (DKFZ), Heidelberg, Germany.

Contents

Introduction	17
1 Medical Image Acquisition	19
1.1 Computed Tomography	19
1.2 Magnetic Resonance	20
2 Convolutional Neural Networks in Biomedical Data	25
2.1 Main Parts of CNN	25
2.2 Observation of U-Net	30
2.3 nnU-Net	30
2.4 Improving Quality of Segmentation	33
3 Experiment	35
3.1 Dataset	35
3.2 Training of Test Model	35
3.3 Analysis of the Training Quality	37
4 Discussion	45
Conclusion	47
Bibliography	49
List of symbols, quantities and abbreviations	53

List of Figures

1.1	Image Contrast	22
1.2	Comparison of CT and MR	23
2.1	U-net Architecture	31
2.2	nnU-Net Architecture	32
2.3	Principal Idea of Similarity	33
2.4	Overview of the workflow of the proposed IDAL-algorithm	34
3.1	Loss Functions of nnU-Net Learning on 2D Dataset	38
3.2	Evaluation Metrics of nnU-Net Learning on 2D Dataset	38
3.3	Examples of Gliomas Segmentation	39
3.4	Similarity Matrix	42
3.5	Dendrogram of Similarity Matrix	43
3.6	Dice Score of Test Images Prediction	44

List of Tables

1.1	MR Signal Characteristic of Different Tissue	21
1.2	Comparison of CT and MR in Neurology	23
3.1	Dataset Description	36
3.2	Comparison of Evaluation Metrics	37
3.3	Changing of Median Dice Score	41

Introduction

Since 2012 computer vision and deep learning techniques and principally convolutional neural networks made a huge breakthrough in image recognition. Until 2012 the sphere of the neural network was developed however not in such way as we have it now. Previously we met two problems: personal computers did not have a suitable capacity (both CPU and GPU) to be used as an instrument for solving the image recognition tasks and the second large problem was the availability of big data. Therefore it was a long way to current outstanding results. The first computer vision department was established in summer 1966 in Massachusetts Institute of Technology, two years before Stanford's school, both of them had a great transformation and development through the current period. Only after several decades, at the beginning of 2010, NVIDIA has presented their powerful-capacity microchips, which were able to process images and as well were open for the mass-market. The issue with big data was covered by the growing project named IMAGENet. Nowadays computer vision and deep learning is a spotlight for many scientists and professionals all over the world as a modern, flexible and multitask tool, which can be used in a wide range of options. As a great example, we can run across the widening number of projects in such international conferences as CVPR and ICCV with more than 2 000 attending researchers each year.

The IMAGENet is a massive dataset with more than 14 million images and 22 thousand categories, which were marked by hand for several years. Convolutional neural networks became popular in 2012 after publications of the results of IMAGENet competition, where the task was correctly to classify and to detect objects and scenes. By some means, it was a revolutionary and historical moment of deep learning. On that year neural network model showed its massive power and high capacity into training architecture with almost half decrease of the error (from 0.26 to 0.16[1]) and continued its winning every single year after it. Since 2012 scientists and enthusiasts from all parts of the world put a lot of effort to develop and widen the sphere and already have archived some great results. The modern convolutional neural networks have a difficult structure and include sometimes more than 1 000 layers. Today we can use it for amount range of purposes: detection, classification, segmentation and not only images but videos and 3D objects can be used too.

Everyday computer technology goes deeper and deeper to our life and no wonder that such techniques already in several years after the breakthrough have come into the biomedical sector. Doctors and researchers are constantly working together to unlock new possibilities to cure or help people through computer vision's power. The first from the range of doctors, who need computer vision and artificial intelligence help, are radiologists. They work with CT scans and MR images, detection of

objects may help them to discover the probable diseases with less error rate, amount of time and in parallel processing. Manual segmentation of the brain tumours in MR images is a time-consuming process, which increases the required time for the research of tumour development. Segmentation of tumours in MR images currently is an achievable task for deep learning algorithms. It is possible to train the CNN model with supervised learning, where the database will include MR images and ground truths with segmenting regions of interests (tumours).

Recently it has already been discovered several fully automatic segmentation algorithms to solve the segmentation problem and to show the importance of medical automatic image recognition. The rise in deep convolutional neural networks performance, due to their abstractions of different levels of features, motivated many researchers to transfer the knowledge acquired by these networks, when trained on millions of images into new tasks such as medical image segmentation, to benefit from their learned parameters, in particular, weights. The most common convolutional neural network for image segmentation task is called U-Net, which issued for biomedical image segmentation at the Computer Science Department of the University of Freiburg, Germany.

The following master's thesis paper equipped with a short description of CT scans and MR images and the main differences between them, explanation of the structure of convolutional neural networks and how they implemented into biomedical image analysis, besides in paper it was tested a popular modification of U-Net with loss-functions. As far as segmentation quality plays a highly important role for doctors, in experiment conclusion part it was paid significant attention to training quality and prediction results of the model.

1 Medical Image Acquisition

The fast advancement of biomedical imaging techniques in the last couple of years has significantly enhanced the role of imaging in biology and medicine. Computer imaging began to play an integral and important role in the assessment of neurological disorders with the evolution of technology. It is accomplished for diagnosis, assessing the efficacy of therapy, follow-up, and guidance for procedures. Benign neurological disorders and life-threatening neoplasms may present with similar overlapping symptomatology, which could be relatively nonspecific. Clinical history combined with a good neurological evaluation is often followed by laboratory investigations, electroencephalography, lumbar puncture, and other investigations. Magnetic resonance (MR) and Computed tomography (CT) form the backbone of the imaging work-up of these patients. Neurological disorders are always tending to be complex and arriving at a diagnosis requires a deep understanding of neuroanatomy, physiology and ability to use last updated diagnostic tools such as neuroimaging.

Imaging Modalities

Previously radiography ('plain films' then digital radiography) significantly affected on medical imaging. With the advent of advanced imaging technologies like CT and MR, the role of radiography has remarkably decreased. Currently, CT and MR are the most commonly performed imaging modalities for the brain and spine diseases. In addition, to extend anatomical details it was developed new advanced MR techniques such as MR perfusion, MR spectroscopy and functional MR (fMRI) provide fuller physiological information. Furthermore, single-photon emission CT (SPECT) and positron emission tomography (PET)-CT imaging have a distinct contribution, as they can provide functional and structural brain images.

1.1 Computed Tomography

CT technology bases on the identical physical principle as x-rays do. The differential absorption of the x-ray beam by different tissues produces varying levels of density in the image, which on CT scans are measured in Hounsfield units. It can be shown in cross-sectional format or multiple planes. Multidetector CT has increased the capability with greater spatial resolution, faster scans and multiplanar reformations[2].

The fast ability of CT to image traumatic conditions of the spine and brain rapidly has made it invaluable in acute neuro operations. Commonly seen lesions on CT include acute intra and extra-axial haemorrhage, hemorrhagic contusions, diffuse axonal injury, spinal and calvarial fractures.

CT plays a crucial and important role in the management of acute stroke and is the first imaging modality due to the quite short procedure time. It works quickly in determining whether the symptoms being observed can be attributed to intracranial haemorrhage, ischemic stroke or a mass lesion. The vital benefaction of non-contrast CT is excluding intracranial haemorrhage, so that appropriately selected patient can be started on tissue plasminogen activator. Even though it is easy to admit that CT is less sensitive than MR in detecting acute cerebral ischemia-infarction, visible and detectable changes are present on nearly 50 per cent of non-contrast CT scans in patients with major territorial infarcts.

Non-contrast CT is relatively insensitive for the detection of neoplastic disease, especially when the tumour is small. A post-contrast CT has to be made when evaluating neoplastic conditions using CT.

1.2 Magnetic Resonance

The MR signal used to generate mostly all clinical images comes from hydrogen nuclei. Hydrogen nuclei are consisting of a single proton which is constantly spinning. A radio frequency pulse (RF pulse) emitted from the scanner results in some of the hydrogen protons and then being knocked in 180° out of alignment with the static magnetic field. The simplest form of the spin-echo (SE) pulse sequence consists of 90° -pulse, a 180° -pulse, and then an echo. The time between the middle of the first RF pulse and the peak of the spin echo is called the echo time (TE). The sequence then repeats at time TR, the repetition time. Because the energy from the RF pulse is dissipated, the hydrogen protons will return to alignment with the static magnetic field. The MR signal is derived from the hydrogen protons as they move back by the magnetic field. The *Magnetic Resonance* (MR) signal is then broken down and spatially placed in resulting images[2].

T1, T2, and proton density (PD) are the basic parameters of MR and determine the contrast between different tissues. Spin and gradient echo are two main sequences in MR. All other sequences are variations of one of these sequences and are used to better characterize specific tissue types. MR sequences that emphasize tissue differences in T1 relaxation are called T1 weighted, and those that emphasize T2 relaxation are called T2 weighted. Mainly tissues with short T1 relaxation time such as fat, protein and melanin produce a high signal on T1-weighted sequences and appear brighter on images, whereas cerebrospinal fluid (*Cerebrospinal Fluid* (CSF)) is relatively darker. Fluids have a long T2 relaxation time and appear bright on T2-weighted sequences. To more close comprehension of imaging properties of the main types of tissues, there is present Tab. 1 above.

Tab. 1.1: MR Signal Characteristic of Different Tissue[2]

	T1	T2
Dense bone	Low signal	Low signal
Fat	High signal	Loses signal compared with T1
Water (cerebrospinal fluid)	Low signal	High signal
Hemorrhage	Variable (depends on stage of hemoglobin breakdown)	Variable (depends on stage of hemoglobin breakdown)
Gray/white matter	White matter higher signal than gray matter	White matter lower signal than gray matter

On spin-echo (SE) imaging, the repetition time (TR) and the echo time (TE) are used to control image contrast and the "weighting" of the MR image. In the broadest sense, T1-weighted and T2-weighted are used to communicate to other physicians the type of MR pulse sequence employed to generate a series of images. In a more narrow sense an implication exists that a single intrinsic tissue parameter (T1, T2, spin-density, diffusion, susceptibility, chemical shift, flow, perfusion, etc.) dominates the image contrast observed. The traditional model in SE imaging to which the main question refers considers four combinations of TR and TE values:

- Short TR/Short TE → T1-W
- Long TR/Short TE → PD-W
- Long TR/Long TE → T2-W
- Short TR/Long TE → not used

The exact reference ranges are not generally specified, but usually "long" TR or TE means 3-5x T1 or T2 respectively, while "short" implies TR or TE \ll T1 or T2.

To understand the effect of TE on T2-weighting is to consider the signals generated by two tissues with different T2 values. When TE is short, the echo occurs when there has been little time for T2-decay to have taken place and hence the tissues are not differentiated. If TE is long, the relative differences in signal decay between the two tissues become more noticeable, and hence more "T2-weighting."

Similar arguments can be made for the interplay between TR and T1. At long TR's tissues with different T1 values have all had time to recover from the 90 excitation pulse, so their signals are not dramatically different. Conversely, short TR's accentuate "T1-weighting".

Finally, when TR is long and TE is short, both T1 and T2 effects are minimized. The only remaining factor is the spin-density [H], which becomes the dominant

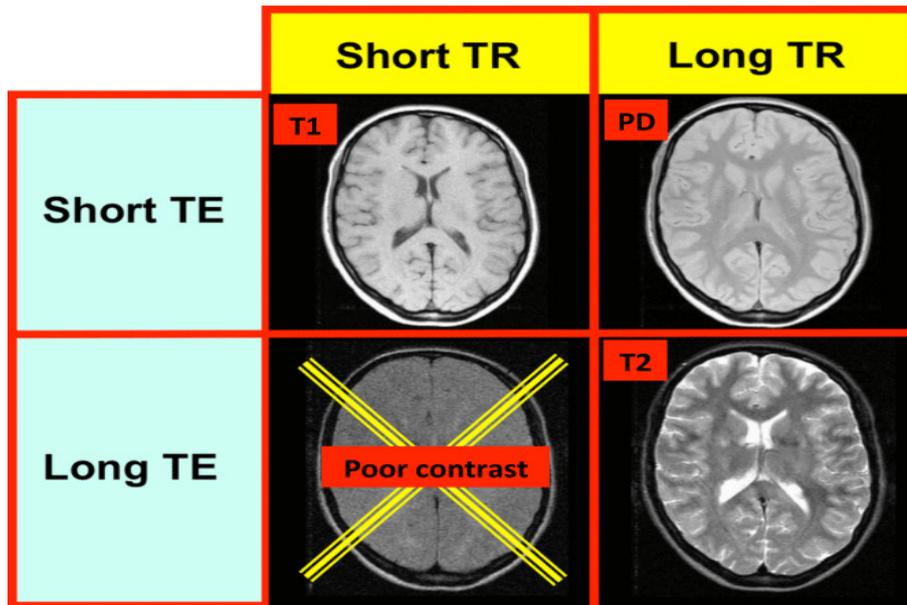


Fig. 1.1: Image Contrast with Different Repetition Time (TR) and Echo Time (TE)[3]

weighting for that combination of parameters. Result of following combinations are described with image examples in the Fig. 1.1 below.

Besides, it may be useful to mention next technique: MPRAGE is a three-dimensional, thin-section T1-weighted volumetric acquisition that is increasingly utilized for evaluating a wider spectrum of brain disorders.

Fluid-attenuation inversion recovery (FLAIR) sequence is used to eliminate the signal from CSF, which comes into view dark. It is useful for highlighting parenchymal lesions that lie close to ventricles or multiple sclerosis plaques or small cortical infarcts, which are not as easily seen on T2-weighted sequences. Bright CSF signal on FLAIR could suggest leptomeningeal disease with the replacement of the normal CSF by blood (subarachnoid hemorrhage), pus (meningitis) or tumour cells (leptomeningeal carcinomatosis).

Summarising the overall analysis of CT scans and MR images it will be useful to look at the table with all the main differences. Tab. 2 was attached above and it shows how the main aspects of tissue appearance on the resulting image, moreover, below it also was attached a practical case in Fig. 1.2 with CT and T2-weighted MR images of leukoaraiosis that shows basic differences in image representation. From the following point of view and previously accumulated experience in medical images, it obviously will be a conclusion that in brain tumour analysis more reasonably to use MR images instead of CT scans. That is why in my master thesis the main dataset includes MR images with following sequences: T1-weighted, T2, *Fluid-attenuation Inversion Recovery* (FLAIR) and T1 Gd.

Tab. 1.2: Comparison of CT and MR in Neurology[2]

Computed tomography	Magnetic resonance
Few contraindications	Multiple contraindications
Fast, readily available	Slow, less available
Radiation exposure	No radiation exposure
Very good for acute hemorrhage	Excellent for different phases of hemorrhage
Poor soft tissue contrast	Excellent soft tissue contrast
Nephrotoxic contrast	Nephrogenic systemic fibrosis
Higher incidence of contrast reaction	Lower incidence of contrast reaction
Need contrast for CTA	Non-contrast MRA (flow related)

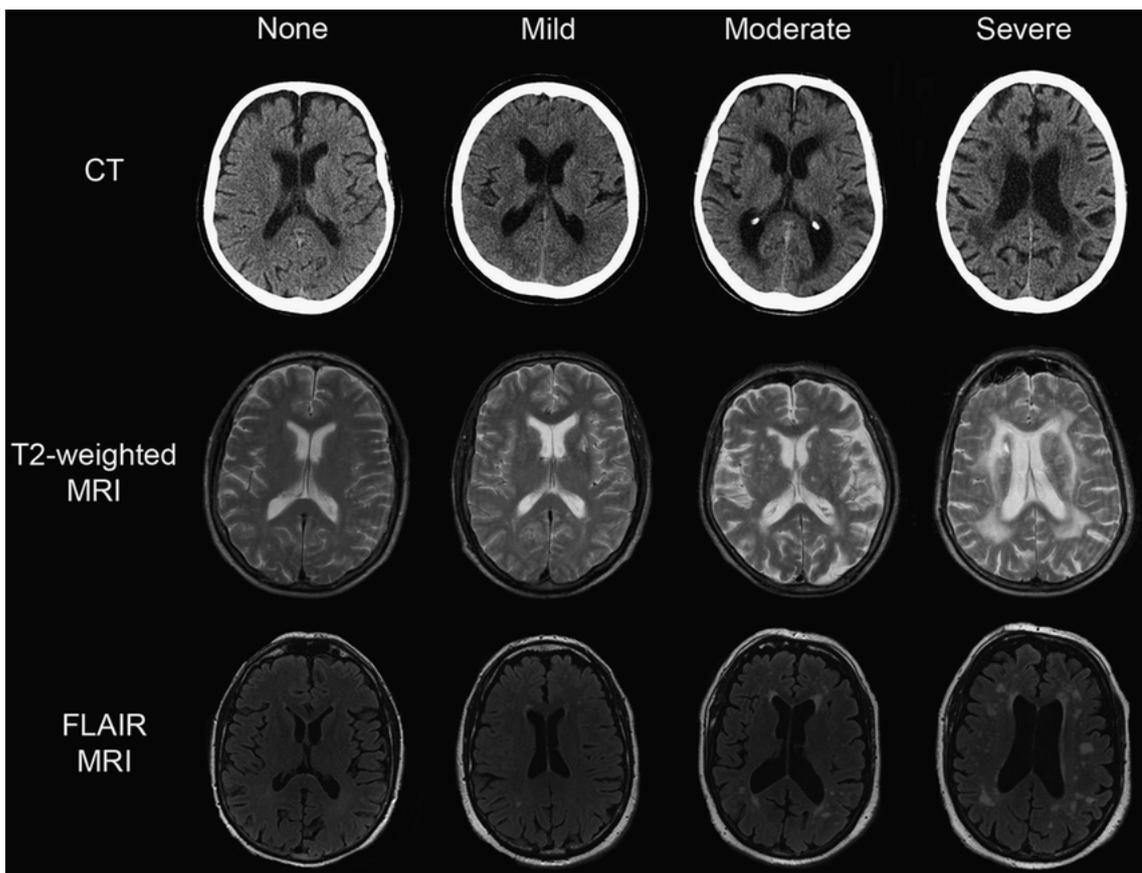


Fig. 1.2: Comparison of CT and MR images on example that shows leukoaraiosis at different stages of severity [4].

2 Convolutional Neural Networks in Biomedical Data

A convolutional neural network is an implementation of a neural network used in machine learning that processes array data such as images, and is thus frequently used in applications aimed at medical images.

A key problem in medical image computing today is that the process of applying a segmentation algorithm to a new problem is completely human-driven. It is based on experience, with the papers mostly focusing on the network architecture, while merely brushing over all the other hyperparameters. Sub-optimal adaptations of a baseline method are regularly compensated for by the proposal of a new architecture. Since the strong dependencies and amount of local minima in hyperparameter space make it really hard to optimally adapt a method to a new problem, nobody in this loop can really be blamed. This situation can be frustrating for the researchers and computer science community. Mainly in the medical imaging domain where datasets are so diverse, progress will largely depend on our ability to solve these problems[5].

2.1 Main Parts of CNN

A typical convolutional neural network is divided into three main parts although sometimes the architectural implementation varies considerably:

1. input (depends on dataset, can be text, image, volume or video)
2. feature extraction (encoder)
3. regression/classification and output

Input of Neural Networks

Input objects of neural networks have many options it can be text, audio, image, video plus in medical application standard signals. In fact of the smaller size to compare with others and less demand for computational power the most common input of CNN currently is an image. Although considerable work has also been performed on 3D convolutional neural networks, which can process either volumetric data (3 spatial dimensions), video (2 spatial dimensions + 1 temporal dimension) or 1D data as signals (biomedical as EEG, EMG, ECG, EOG, DNA sequences, etc.).

Mostly in implementations, the input needs to be processed to match the particulars of the *Convolutional Neural Network* (CNN) being used. This may include reducing the size of the image, cropping, identification of a particular region of interest (ROI), as well as normalizing pixel values to particular regions.

Feature Extraction (Encoder) Part

The feature extraction part is what tells the difference between CNNs from other multilayered neural networks. It typically comprises of repeating sets of following sequential steps:

- convolution layer
 - input is convoluted by application of numerous kernels
 - every kernel results in a feature map
- non-linear activation unit

In neural networks, the activation function of a node defines the output of that node given an input or set of inputs. Experience shows that only nonlinear activation functions allow such networks to compute nontrivial problems using only a small number of nodes[6].

- the activation of each neuron is then computed by the application of a non-linear function to the weighted sum of its inputs and an additional bias term. It gives the neural network the ability to approximate almost any function.
- currently, well-spread an activation unit called the Rectified Linear Unit (ReLU)[7] and their modifications: Bipolar rectified linear unit (BReLU)[8], Leaky rectified linear unit (Leaky ReLU)[9], Randomized leaky rectified linear unit (RReLU)[10], Exponential linear unit (ELU)[11] and others.
 - * during convolution and pooling processes results in some pixels are having negative values
 - * the rectified linear unit verifies all negative values are at a zero
- pooling layer every feature map is downsampled into a smaller matrix by pooling the values in neighboring pixels

These steps run through again certain amount of times, each convolution layer acting upon the pooled and rectified feature maps from the previous layer. The result is an ever-smaller matrix size with activation dependent on growing complex features due to the cumulative interaction of numerous prior convolutions.

Regression/Classification and Output Part

Most frequently convolutional neural networks in radiology undergo supervised learning. It involves training an algorithm from a set of images or data where the output labels are already known[12]. Supervised learning is broken into two subcategories, classification and regression[13]. Classification refers to the prediction of whether an image falls into one or more categories while regression aims to predict a continuous label. The final pooled and rectified feature maps are used as the input of fully connected layers just like in a fully connected neural network.

Regression and classification are categorized under the same umbrella of supervised machine learning. Both share the same concept of utilizing known datasets (referred to as training datasets) to make predictions. In supervised learning, an

algorithm is employed to learn the mapping function from the input variable (x) to the output variable (y); that is $y = f(X)$. The objective of such a problem is to approximate the mapping function (f) as accurately as possible such that whenever there is a new input data (x), the output variable (y) for the dataset can be predicted. The main difference between them is that the output variable in regression is numerical (or continuous) while that for classification is categorical (or discrete)[14].

Last layer in regression is usually fully connected, it does not have activation function (sometimes in the specific cases ReLu or basic activation function building block Softmax[15]), number of neurons corresponds to the number of regressed values.

Otherwise, last layer in classification is fully connected too but frequently with softmax (sigmoid) activation function and number of neurons corresponds to the number of classes. Output vector in classification includes values of membership probabilities for each class (length of the vector equals the number of classes). Resulting estimated classes is taken with maximal probabilities.

Training of Neural Network

As it was mentioned in the introduction part, the most popular and wide-spread convolutional neural network for segmentation purposes is called U-Net[16]. There are several other options for segmentation task like Medical Detection Toolkit[17] or Mask R-CNN[18], however, with more detailed look, we probably may see that other options have mostly similar structure or even include U-Net in themselves. From this perspective let us make an observation of the most popular and powerful instrument for biomedical image segmentation.

Training of the deep model includes direct work with optimization and loss functions and if we are talking about optimizers researches already achieved some high-rated results but loss functions choice sometimes maybe only as an experiment results or a guess.

Fundamentally, the optimization of supervised machine learning consists of two phases that are looped through continuously:

- forward propagation: for a given input, calculates a predicted outcome and compares this with the expected outcome to give an overall cost (error),
- backward propagation: from the cost, works backwards through the network updating the parameters in an attempt to minimize the overall cost[19].

There are many successful options optimizers for training neural network - gradients based, evolution algorithms, etc. However, recently there were developed a lot of types of gradient based methods (SGD[20], SGD with momentum[21], Nesterov accelerated gradient[22], AdaGrad[23], Adam[24], etc.). U-Net and many others its modifications use (ex. nnUNet[5]) Adam optimizer.

Adaptive Moment Estimation (Adam)[25] is method that computes adaptive learning rates for each parameter. In addition to storing an exponentially decaying average of past squared gradients $v(t)$, Adam keeps an exponentially decaying average of past gradients $m(t)$, similar to momentum. Whereas momentum can be seen as a ball running down a slope, Adam behaves like a heavy ball with friction, which thus prefers flat minima in the error surface[26]. We compute the decaying averages of past and past squared gradients $m(t)$ and $v(t)$ respectively as follows:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (2.1)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (2.2)$$

$m(t)$ and $v(t)$ are estimates of the first moment (the mean) and the second moment (the uncentered variance) of the gradients respectively, hence the name of the method. As $m(t)$ and $v(t)$ are initialized as vectors of 0's, the authors of Adam observe that they are biased towards zero, especially during the initial time steps, and especially when the decay rates are small (i.e. β_1 and β_2 are close to 1).[27]

They counteract these biases by computing bias-corrected first and second moment estimate:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (2.3)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (2.4)$$

They then use these to update the parameters just as we have seen in Adadelta and RMSprop, which yields the Adam update rule:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t \quad (2.5)$$

The authors propose default values of 0.9 for β_1 , 0.999 for β_2 and 10^{-8} for ϵ . They show empirically that Adam works well in practice and compares favorably to other adaptive learning-method algorithms.

Adam optimizer is using with an initial learning rate of 3×10^{-4} for all experiments. Whenever it did not improve by at least 5×10^{-3} within the last 30 epochs, the learning rate was reduced by factor 5. The training was terminated automatically if it did not improve by more than 5×10^{-3} within the last 60 epochs, but not before the learning rate was smaller than 10^{-6} [28].

Several different multi-class loss functions are using for segmentation tasks and currently it is among the main aspects of a successful model. It can be Cross-Entropy (CE), Weighted Cross-Entropy (WCE), Soft Dice[29], Batch Soft Dice,

Tversky Loss[30], Lovász-Softmax[31] or combinations of them. One of the main challenges with brain tumour segmentation is the class imbalance in the dataset. While networks will train with cross-entropy loss function, the resulting segmentations may not be ideal in the sense of the Dice Score they obtain. Since the Dice Scores are one of the most important metrics based upon which contributions are ranked, it is imperative to evaluate this metric. To demonstrate the effectiveness of nnU-Net we can compare results of training with different loss functions: popular cross-entropy and a unique combination of dice and cross-entropy loss from MIC-DKFZ team. The experimental part and results are shown in the section 3.2 Training of Test Model.

Let the number of image patches x_i in our training mini-batches be I and let each image patch consist of C pixels. The segmentation model then maps each of $I \times C = N$ pixels in the mini-batch to probability p_l for each of labels L . The training procedure ensures that the resulting output label probability vectors p_l^c correspond to one-hot encoded ground truth label vectors r_l^c as best as possible on the training data.

Cross-Entropy Loss. Also known as log-loss, cross-entropy is the most widely used loss function for classification CNN. When applied to a segmentation task, cross-entropy measures the divergence of the predicted probability from the ground truth label for each pixel separately and then averages the value over all pixels in the mini-batch:

$$L_{CE} = -\frac{1}{N} \sum_{c=1}^N \sum_{l=1}^L r_l^c \log(p_l^c) \quad (2.6)$$

This loss function tends to under-estimate the prediction probabilities for classes that are under-represented in the mini-batch which is inevitable in training data.

Combination of Dice and Cross-Entropy Loss. By default models in nnU-Net are trained from scratch and evaluated using five-fold cross-validation on the training set. The network trains with a combination of dice and cross-entropy loss:

$$L_{total} = L_{dice} + L_{CE}[28] \quad (2.7)$$

For 3D U-Nets operating on nearly entire patients (first stage of the U-Net Cascade and 3D U-Net if no cascade is necessary) we compute the dice loss for each sample in the batch and average over the batch. For all other networks it is interpreted the samples in the batch as a pseudo-volume and compute the dice loss over all voxels in the batch. Based on past experience, the team of developers is implemented the dice loss as follows:

$$L_{dice} = -\frac{2}{|K|} \sum_{k \in K} \frac{\sum_{i \in I} u_i^k v_i^k}{\sum_{i \in I} u_i^k + \sum_{i \in I} v_i^k} [28], \quad (2.8)$$

where u is the softmax output of the network and v is a one hot encoding of the ground truth segmentation map. Both u and v have shape $I \times K$ with $i \in I$ being the number of pixels in the training patch/batch and $k \in K$ being the classes[28].

2.2 Observation of U-Net

The U-Net architecture is built as the fully convolutional network and modified in a way to produce better segmentation in the sphere of medical imaging. In Fig. 2.1 below, it is shown the visual structure of U-Net and it can be described as comprising of two parts an encoder (contraction path) on the left side and a decoder (expansion path) in the right side. The architecture of the network is noteworthy and reminds the letter "U" because of it the network got such name "U-Net"[16].

The contraction path includes a repeated application of two 3x3 convolutions, where each followed by a ReLU and a 2x2 max pooling operation with stride 2 for downsampling. At every downsampling step, the number of feature channels is doubling. This captures context via a compact feature map[16].

The expansion path includes upsampling of the feature map followed by a 2x2 convolution that halves the number of feature channels a concatenation with the cropped feature map from the contracting path, and a 3x3 convolutions, followed by a ReLU. The upsampling of the feature dimension is done to meet the same size as the block to be concatenated on the left. The expansion increases the “what” which helps in getting more features but losses the localization, localization information is concatenated from the contraction path. The cropping is necessary due to the loss of border pixels in every convolution. At the final layer, a 1x1 convolution is used to map each 64-components feature vector to the desired number of classes[16].

Since 2015 the U-net implementation has achieved proficient performance on different biomedical segmentation applications. It needs a small number of ground truths and has an outstanding reasonable training time. All these qualities make U-Net is a perfect candidature as a tool in the hands of many researchers for tumour segmentation.

2.3 nnU-Net

As it was described above U-Net[16] now is the state-of-art technique for image segmentation. However, U-Net is modified by several groups of scientists with some up-

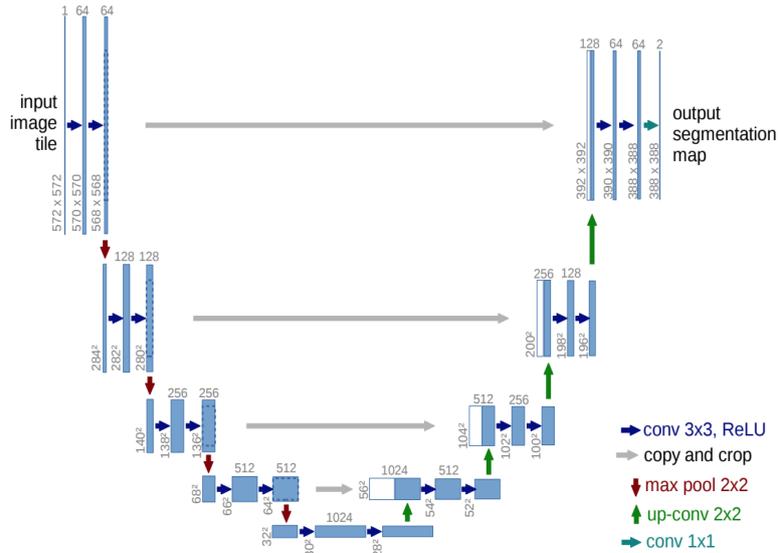


Fig. 2.1: U-net Architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations[16].

grades for each specific aim. In the case of brain imaging sphere, great achievements are shown in the Multimodal Brain Tumor Segmentation Challenge (BRATS)[32] provided by Center for Biomedical Image Computing and Analytics (CBICA) and supported by non-profit corporation MICCAI. These companies organize challenge each year and publish diverse datasets with ground truths for two tasks: segmentation and survival. For us, the segmentation task has a significant interest. From the rankings, we can find the best result techniques and work on it because usually, they are in open source.

From all teams in the BRATS competition there is one — MIC-DKFZ, which demonstrated the effectiveness of a well trained U-Net and was in Top 3 during last three years. It is Division of Medical Image Computing in German Cancer Research Center (DKFZ) from Heidelberg, Germany. Team of scientists: Fabian Isensee, Philipp Kickingereder, Wolfgang Wick, Martin Bendszus and Klaus H. Maier-Hein proposed their own version of U-Net: No New-Net or just nnU-Net[33]. Due to the sheer number of such variants, it becomes increasingly difficult for researchers to keep track of which modifications extend their usefulness over the few datasets they are typically demonstrated on. Unlike other segmentation methods published recently, nnU-Net does not use complicated architectural modifications and instead revolves around the popular U-Net architecture. MIC-DKFZ team have implemented a num-

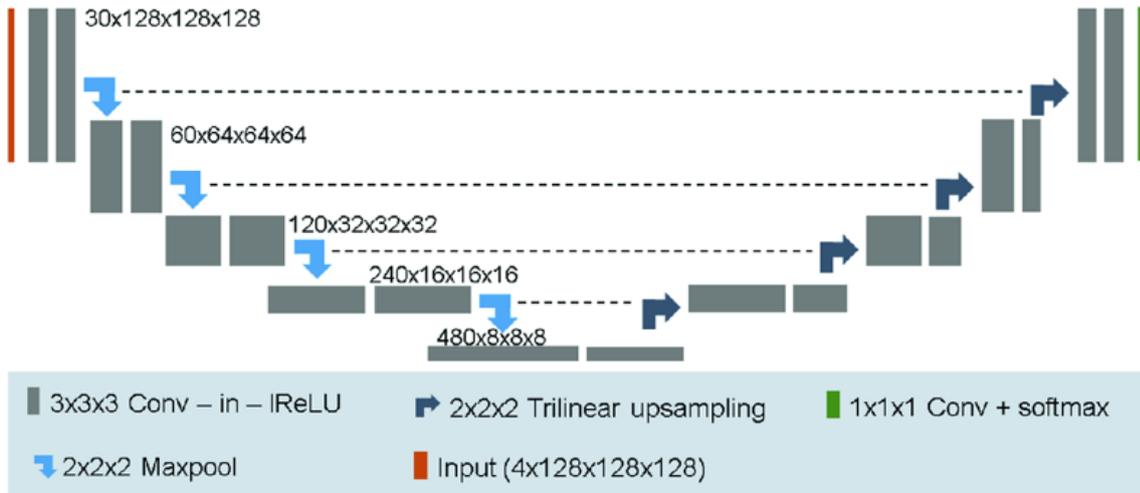


Fig. 2.2: nnU-Net is a 3D U-Net architecture with minor modifications. It uses instance normalization and leaky ReLU nonlinearities and reduces the number of feature maps before upsampling. Feature map dimensionality is noted next to the convolutional blocks, with the first number being the number of feature channels[33].

ber of these variants and found that they provide no additional benefit if integrated into a well trained U-Net. In this context, contribution to the BRATS challenges was intended to demonstrate that such a U-Net, without using significant architectural alterations, is capable of generating competitive state-of-the-art segmentations.

The architecture of nnU-Net is an instantiation of the 3D U-Net with minor modifications. It was stuck with design choice to process patches of size $128 \times 128 \times 128$ with a batch size of two. Due to the high memory consumption of 3D convolutions with large patch sizes, the network was implemented carefully to still allow for an adequate number of feature maps. By reducing the number of filters right before upsampling and by using in-place operations whenever possible, this results in a network with 30 feature channels at the highest resolution. Due to previous experience of the team, traditional ReLU activation loss function did not reliably produce the desired result that is why it was replaced with leaky ReLUs (leakiness 0,01) throughout the entire network[33]. With a small batch size of 2, the exponential moving averages of mean and variance within a batch learned by batch normalization[34] are unstable and do not reflect the feature map activations at test time very well. Instance normalization is providing more consistent results and therefore used it to normalize all feature map activations (between convolution and nonlinearity). Overview of architecture is shown in Fig. 2.2.

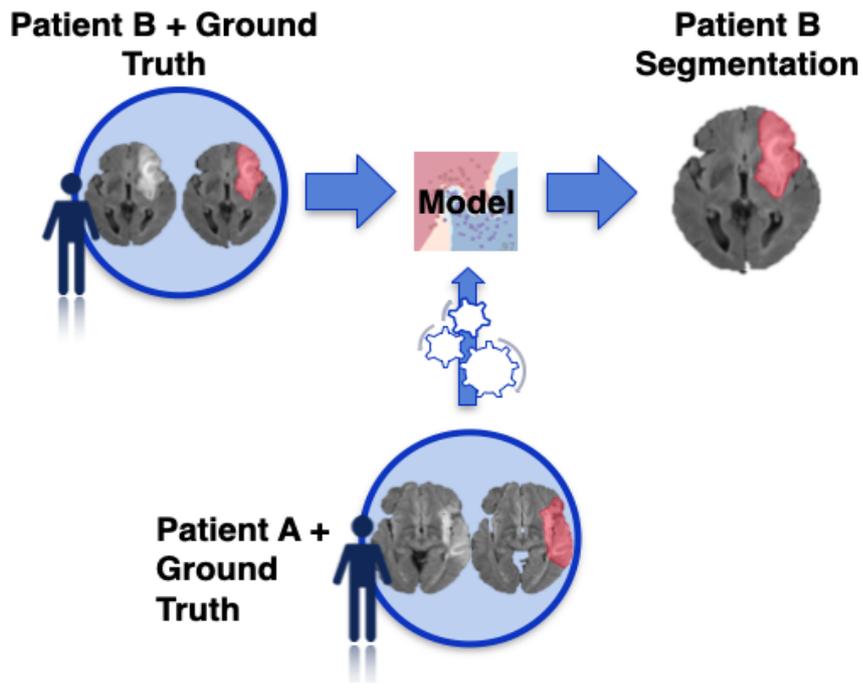


Fig. 2.3: Principal idea of similarity. Similarity of Patient A to Patient B is defined as the quality of the automatic segmentation of Patient B from a classifier trained on Patient A

2.4 Improving Quality of Segmentation

The quality of automatic segmentation still is important issue for doctors, who are planning to use deep learning techniques their practice. The final and ideal task of the thesis is to verify and improve the quality of image segmentation in convolutional neural networks and optimize the training dataset through its similarity.

There are several ideas on how to optimize the quality of segmentation can be found. After some research, the idea of using similarity and 'Input Data Adapted Learning' (IDAL)[35] seems promising for the case where we already have some good segmentation result and would like to improve it. It proposes to learn the best training base for every image and use this to predict a subgroup of best training images for every previously unseen image. It has shown potential with Random Trees techniques, however, the method may be treated universally on other deep learning techniques including convolutional neural networks, because it is based on the similarity between Patient A and Patient B and proposes a concept how to find and fit a special classifier — Similarity Classifier (SC), which optimizes the best dataset for Patient B, look at Fig.2.3. As it was mentioned earlier, a similar experiment carried out only on Random Trees and never on CNN, which give the following thesis an innovation.

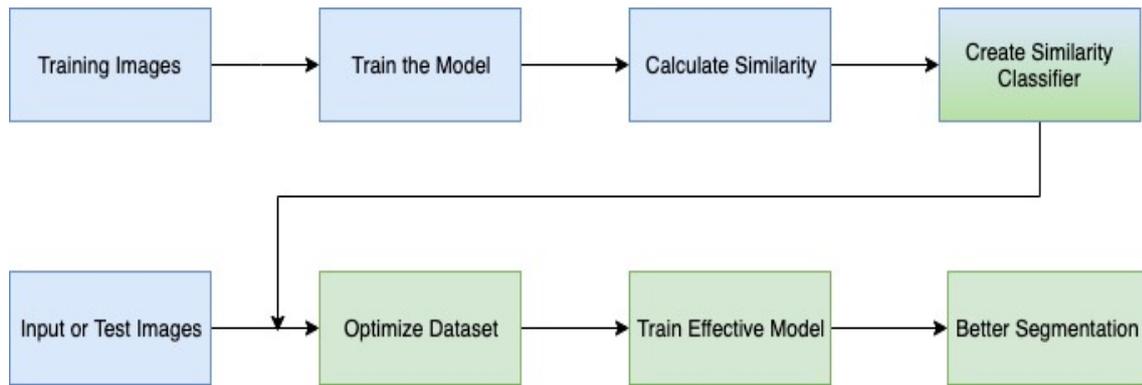


Fig. 2.4: Overview of the workflow of the proposed IDAL-algorithm

Instead of training a single classifier that is used to predict all unseen images method proposes to adaptively train a new classifier for every new image. This allows us to use only few, but similar images during the training. While such an approach makes each classifier less general, it is expected that the so-trained classifier is better suited to deal with the afore mentioned heterogeneity.[35]

This approach with a two-staged algorithm (Fig.2.4). During the first stage, that is performed classifier training, we create a similarity classifier (SC) which can group images based on some similarity measures. In the second stage to find images that are similar to the new, unlabeled image or just similar types of images (ex. liver, brain etc. tumours). From the new optimized dataset, we can train more effective model with better qualitative segmentation results.

3 Experiment

3.1 Dataset

As it was mentioned before one of the biggest open-source brain image datasets is provided by BRATS[32] competition. The vast majority of nnU-Net[33] was developed and organized in the context of the Medical Segmentation Decathlon[36] among different tasks there is brain tumour. The main attributes of the brain tumour dataset from are collected the Medical Segmentation Decathlon in Tab.3.1. The whole dataset is quite massive for training but well-build to prevent overfitting, which makes a good balance. It consists of two years of BRATS competition: 2017 and 2018 and together has 484 training + 266 testing images and with total size near 9Gb. Moreover, on the stage of preprocessing dataset expands dramatically in the size and requires at least 20 Gb on the RAM disk.

NnU-Net uses three types of U-Net models and can automatically choose which (of what ensemble) of them to use. The default setting in a five-fold cross-validation is to train each of these models:

- 2D U-Net
- 3D U-Net (full resolution or fuller)
- 3D U-Net Cascade

From the authors, it is known that 3D U-Net model (3d fullres) looks like the most robust way. If the patch size of the 3D U-Net only covers a very small fraction of an image then it is possible that the 3D U-Net cannot capture sufficient contextual information in order to be effective. If this is the case, it would be better to consider running the 3D U-Net cascade (3d lowres followed by 3d cascade fullres). If data is very anisotropic then a 2D U-Net may be a better option[5]. Also, it will be useful to take into account that each next type of running requires unique preprocessing of the data and more RAM space on the disc compare to the previous one.

3.2 Training of Test Model

To begin training model I have to find the GPU power. It can be realized in several ways: personal computer, supercomputer or virtual server/environment. Due to no opportunity to use personal or supercomputer on the beginning stage of thesis, it was chosen virtual environment with free access to GPU. Currently, it is a few options for free GPU: kaggle and Google Colab. Google Colab offers best conditions: 37 GB on disc space, Tesla K80 GPU and intuitively understandable environment based on Jupyter Notebook. These characteristics should be sufficient for preliminary evaluation of the network and even for some experiments.

Tab. 3.1: Brain Tumours Dataset Description[36]

Target:	Gliomas segmentation necrotic/active tumour and oedema
Modality:	Multimodal multisite MRI data (FLAIR, T1w, T1gd, T2w)
Size:	750 4D volumes (484 Training + 266 Testing)
Source:	BRATS 2016 and 2017 datasets
Challenge:	Complex and heterogeneously-located targets

The code of the nnU-Net with description is in open access and free to load. To provide a robust experiment it should be done some preparation and test experiment and to understand and solve some issues, which can appear during the process of the experiment, besides, we can receive the training weights, which can be used as initials in the full experiment. For that task, it was chosen a test of Cross-Entropy (Equation 2.6) and original Combination of Dice and Cross-Entropy (Equation 2.7) loss functions. The first problem has shown when the 3D U-Net with a full dataset on Google Colab has been on a trial run of preprocessing and unpacking the model stage. All the available size on the disc space was occupied by the image data and there was no free space for training weights. To operate the model properly the size of the dataset has been reduced from 484 training images to 100 and 100 for testing.

After trimming of image dataset to a suitable size for Google Colab it was loaded into the system. When preprocessing is done, training of the model can begin. Another problem with limited GPU power has appeared when the 3D U-Net began running. By the reason of 12 hours limit of processing in Colab, the model does not have time to be adequately processed and to reach the maximum of capacity. The learning has been moved slowly with overestimated computational time and nearly 2 000 seconds for each epoch. The learning was stopped on 16 epochs because it took a long time. The only way to increase the speed of computation is to choose a less powerful model type — 2D. The resulting number of the loss function is presented in Tab.3.2.

To prevent the long and time-consuming process in the second attempt, the model was training on 2D and with cross-entropy loss function. The calculation time of each epoch took almost 750 seconds, which was allowing us to calculate 30 epochs and to stop it by order due to slow changes in the loss function during last 15 epochs. Analyzing the results of test in Fig.3.1, it shows us the good learning rate of loss function, which is close in the end to 0 (minimum) and can be assessed and used as the main loss function for many image segmentation tasks.

On the third attempt, the model received the combination of dice and cross-entropy loss function on 2D data. The calculated numbers of function are negative because minimum of the loss function lays on point -1. One epoch consumed lesser

time (650 seconds) and shows as well the appropriate loss function curve in Fig.3.1.

From this test, we may lead to the confirmation that both loss functions: cross-entropy and combination of dice and cross-entropy loss are suitable for image segmentation of brain tumours and can be used in a process of training of the convolutional neural network. However, if we want to compare them and learning rate, we have to use another parameter - Evaluation Metric. In my case I was using Dice Score(F1). From Fig.3.2, model with combination of dice and cross-entropy loss function on 2D data is considering for a leading result because of the slightly bigger number of evaluation metric. The final values of the tested loss functions collected in Tab.3.2.

To display nnU-Net segmentation results we can perform files (predictions) in standard bioimage *.nifty format. After we overlap it with a test image and we are able to see the predicted place of gliomas in the brain. In Fig.3.3 there are two examples from tested images. However, the most important from the technical processing side outcome is training weights. These training weights storage the values, which already can successfully predict the brain tumour, and the highest result will be used as pre-training weights on the next step in quality analysis.

Tab. 3.2: Comparison of Evaluation Metrics (Dice Score(F1))

Training Model	Evaluation metric (F1)
3D fullers dataset with combination of dice and cross-entropy loss (16 epochs)	0.7951
2D with cross-entropy loss (30 epochs)	0.8476
2D with combination of dice and cross-entropy loss (30 epochs)	0.8684

3.3 Analysis of the Training Quality

For analysis of the training quality between several cases, we will use recorded modalities of one patient for each training case consequently. To make the experiment more detailed we took again 100 training images and 100 test patients images from BRATS dataset. Then we compare predicted segmentations with hand-marked ground truths by doctors from the original dataset, evaluate them and cluster or classify them as in was described in Fig.2.4. The presented similarity method can

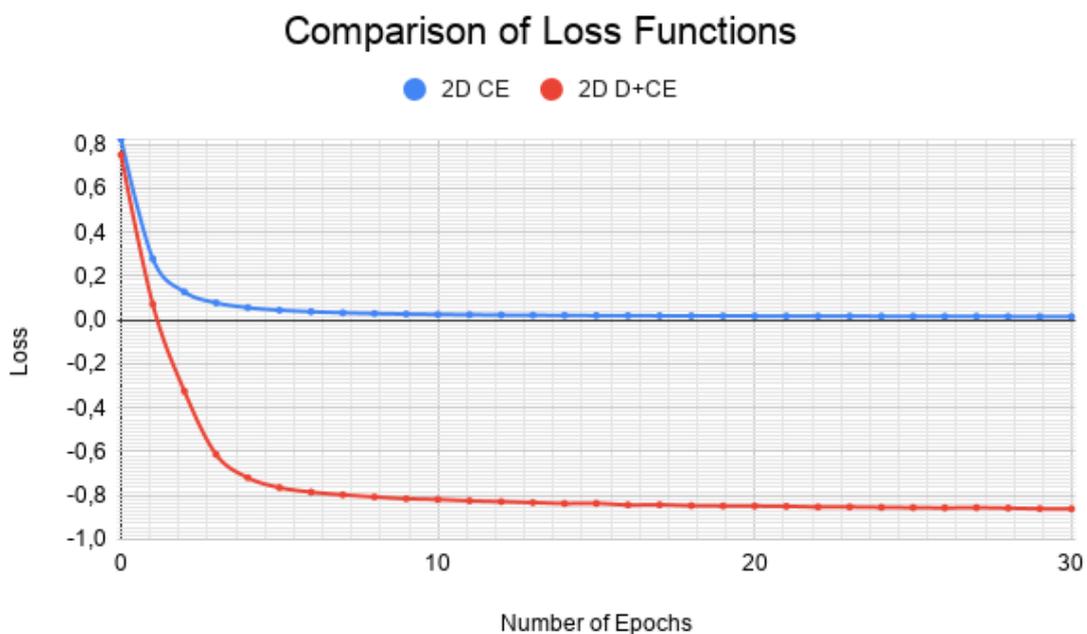


Fig. 3.1: Loss functions values of nnU-Net learning on 2D dataset with cross-entropy and combination of dice and cross-entropy.

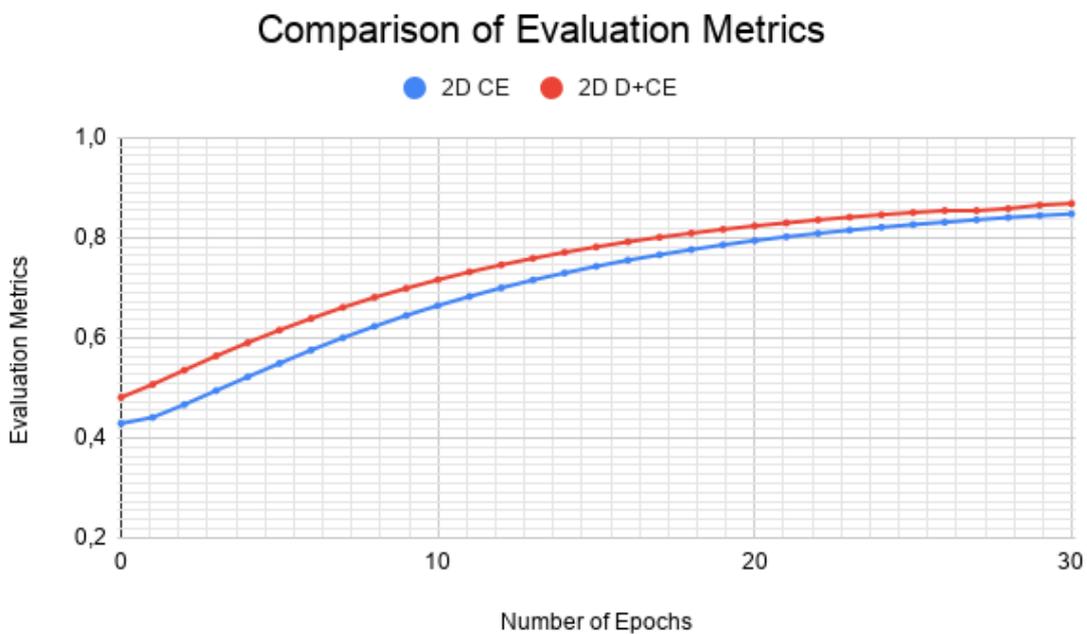
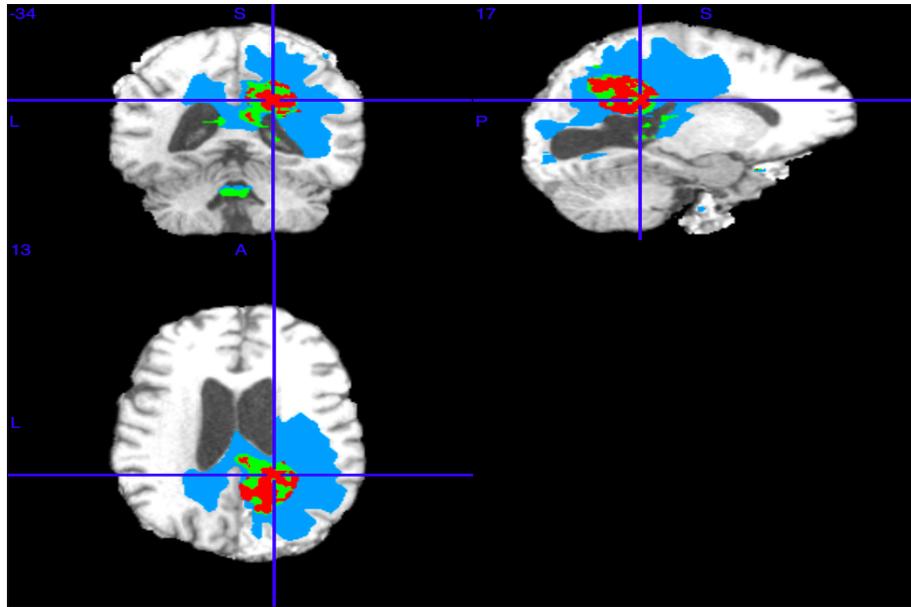
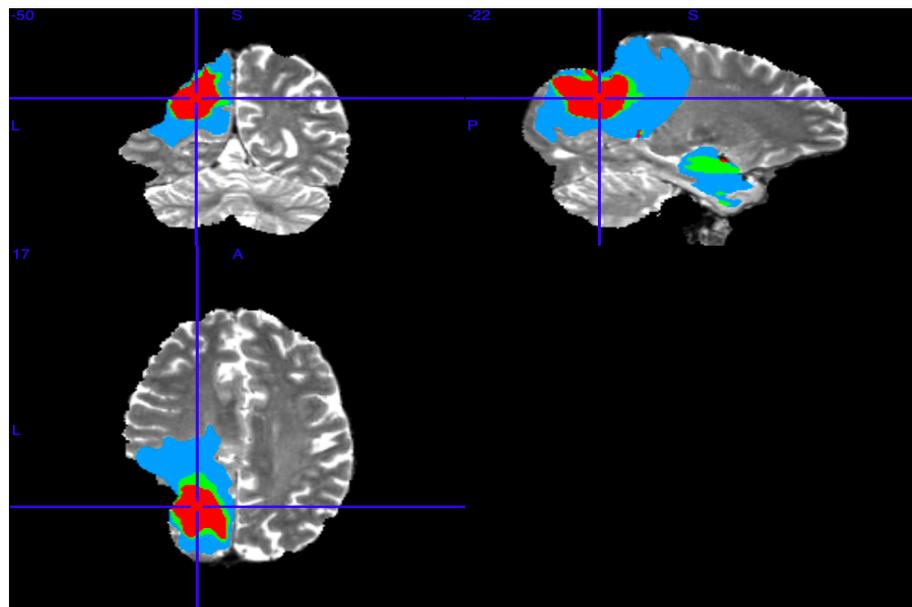


Fig. 3.2: Evaluation metrics of nnU-Net learning on 2D dataset with cross-entropy and combination of dice and cross-entropy loss functions.



(a) Example 1. Results of a Single Slice for Test-patient 05, T1w scan



(b) Example 2. Results of a Single slice for Test-patient 06, T2w scan

Fig. 3.3: Examples of Gliomas Segmentation on the initial training weights, where blue colour is edema, green — non-enhancing tumor and red — enhancing tumour.

be used only, if the ground truth is known, which makes it unusable in real circumstances. However, if the final aim is to achieve a better quality of predictions in image segmentation, it can be implemented via optimized training dataset. Optimizing of the dataset will be organized with similarity classifier that needs ground truths only on the first analysis stage. To generate the automatic classifier, which can evaluate and optimize image datasets and also will be universal for all types of similar images.

The main experiment includes four steps. In the first step, it was a preparation of training weights. For such purpose, we take the pre-training weights from the previous test, where the network has been trained by 30 epochs on 100 images with the most effective result — 2D with combination of dice and cross-entropy loss function. The obtained model was saved with its training weights. The saved weights will be used as initial conditions in the next steps. There was made due to the reason for preventing fail and also better visualizing the further training results with the possibility to give them assessment as upgrade or downgrade. Besides, from the received model, we can predict segmentation and it can give us comparison before and after implementing the classifier.

The second step consisted of preparation and setting up new input of training. Unfortunately, nnU-Net does not allow training the model on only one image. To complete the requirements we used a data augmentation process — batchgenerators by MIC, DKFZ[37], wherefrom one original image converted into a set of five with random mirroring by X, Y, Z axes and spatial deformation with the low scale of the deformation parameters.

In the next step, we were training 100 separate models with obtained previously modalities for 100 extra epochs. When the training process is done, we can perform inference of brain tumour segmentation for chosen dataset — 100 test images. The calculation of segmentation quality will be provided through the similarity between predicted ground truth and the original one. For comparison of segmentation usually, it is used intersection-over-union (Jaccard Index or just IoU) or Dice Coefficient (F1 Score)[38]. Both of them are very similar. The Dice Coefficient calculates from the equation 3.1:

$$Dice = (2 \times Area\ of\ Overlap) / (Total\ Pixels\ Combined) \quad (3.1)$$

As a result, we have a value in the range from 0 to 1, where 1 signifies the greatest similarity between predicted and truth. Collecting all the values in a table and visualizing them on the heatmap can get us a similarity matrix (Fig.3.4), i.e. one cell of the matrix corresponds to the Dice Score, that is trained on the corresponding

training image, is used to segment the corresponding test image[35]. The resulting matrix indicates how similar each patient training case to each other.

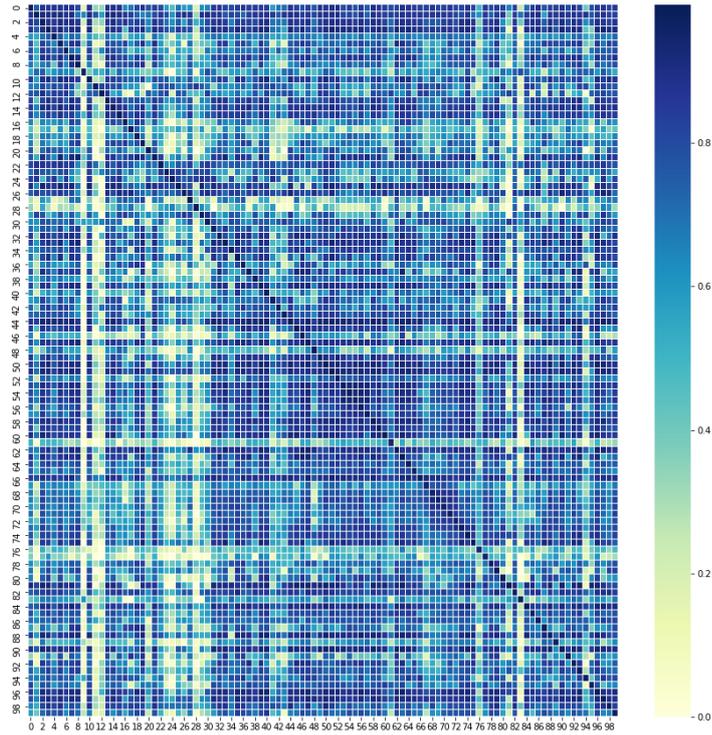
With the last step of experiment from the heatmap we built a dendrogram with the average method (also called the UPGMA algorithm[39]) for analyzing the results. The following dendrogram (Fig.3.5) gives us several clusters of images, where each of them shows us a selection sample for training certain case.

To prove previous statement, we can remove some images or group of images (clusters) that show bad learning rate from dataset. It gives us a new reduced dataset and after training new model on that dataset we can compare predictions of the test images. Dendrogram sorted bad learning cases on the right side on the graph. If we take the worst images, which have the brightest colours, it will be test patients data numbers 12, 11, 83, 9, 28, 24, 81, 80 and 42 from Fig.3.5. To follow the aim to create several examples on the training weights of original 100 training images it was made a prediction of 100 test images and results were evaluated with original ground truths by Dice Score metric and collected in Fig.3.6 a, then dataset was reduced with the first 5 images from the list of worst cases, trained and evaluated. The identical process has been repeated from 7 and 9 images and results are consequently presented in Fig.3.6 a.

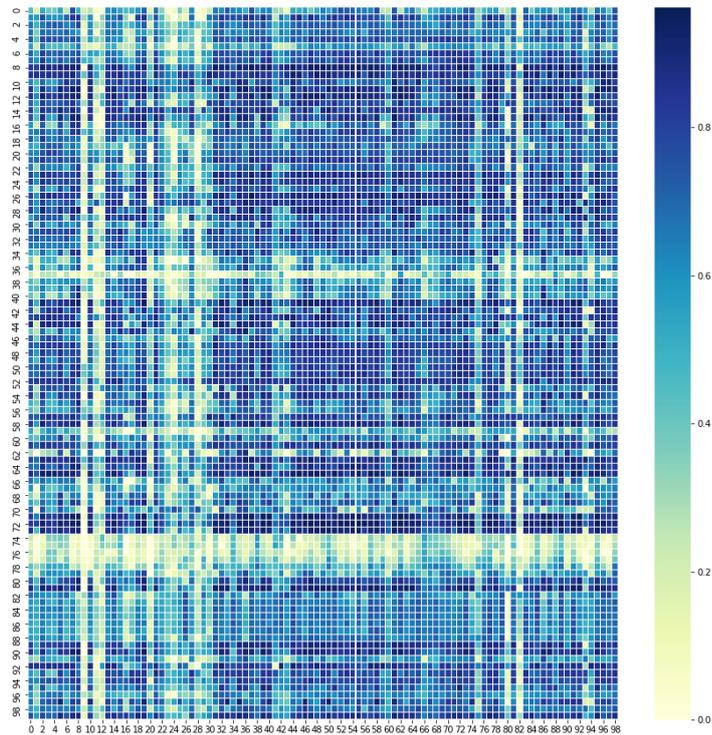
Making the results more clear and visible I subtracted from the original 100 images dataset Dice Score our new dataset Dice Score and collected them in Fig.3.6 b. From that graph, it is possible to say that a few predictions lost up to 15% but in general many images showed reverse effect and gained up to 5% of their quality. Giving assessment to all images it was calculated total median result Dice Score consequently and gathered in Tab.3.3. It has show improvement of only in the case were 7 images were removed and it is already 4.8% of quality. That fact we can achieve better quality from Input Data Adapted Learning method from images similarity has proven. Moreover, the reduced dataset and less computation time compare with the first calculation give us belief that turning dataset in several ways we can reach even more improvement of quality with fewer computation expenses in future.

Tab. 3.3: Changing of median Dice Score relatively original 100 images dataset in shares

Amount of Images in Dataset	Median Dice Score
95	-0.007212
93	0.048353
91	-0.013250



(a) Similarity Matrix of Training Images



(b) Similarity Matrix of Test Images

Fig. 3.4: Similarity matrices, where one cell of the matrix corresponds consequently to the Dice Score of training and test images, that is trained on the corresponding training image. As deeper blue colour (correspond to 1) we have then more similarity between produced segmentation and hand marked ground truth.

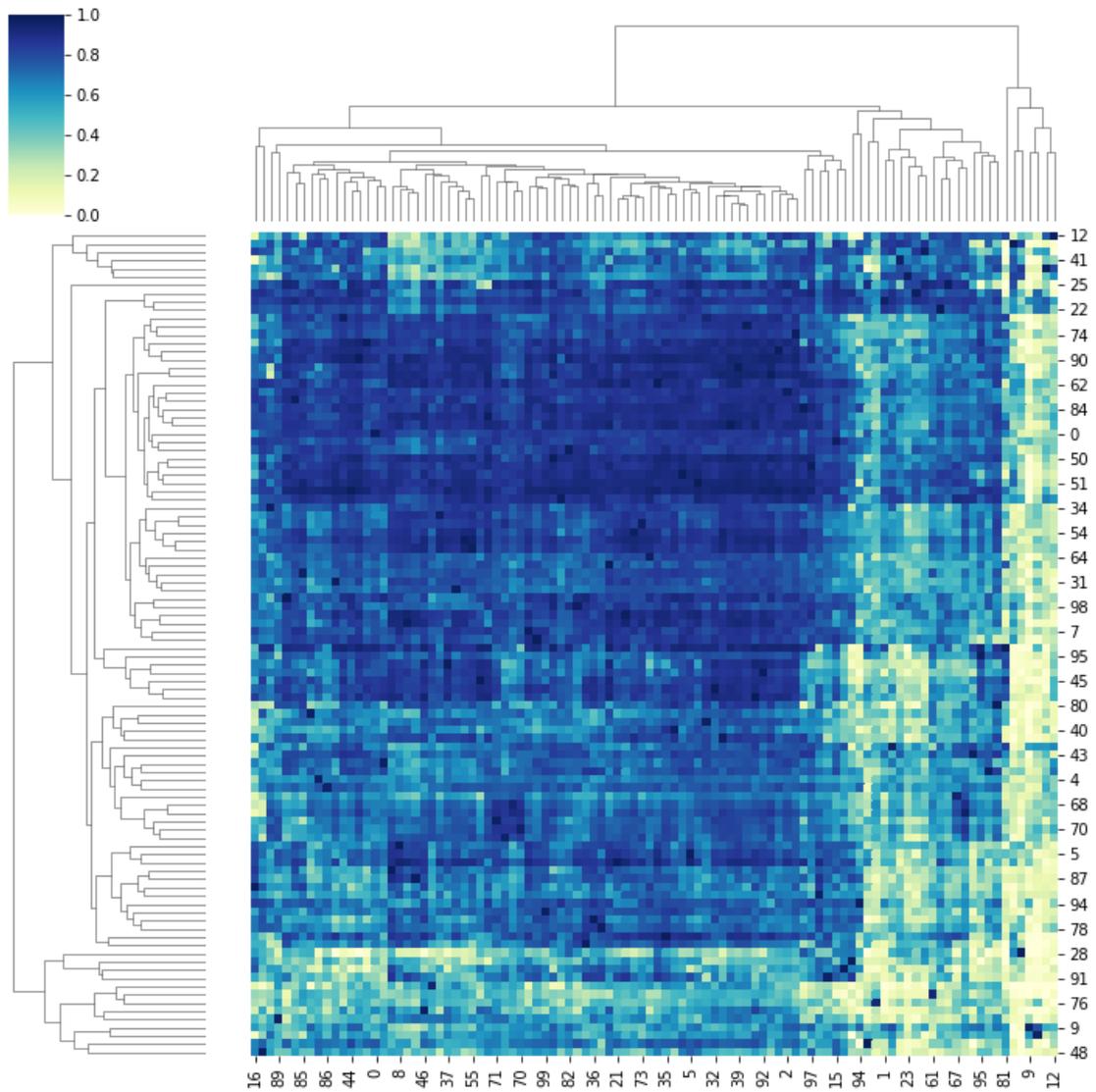
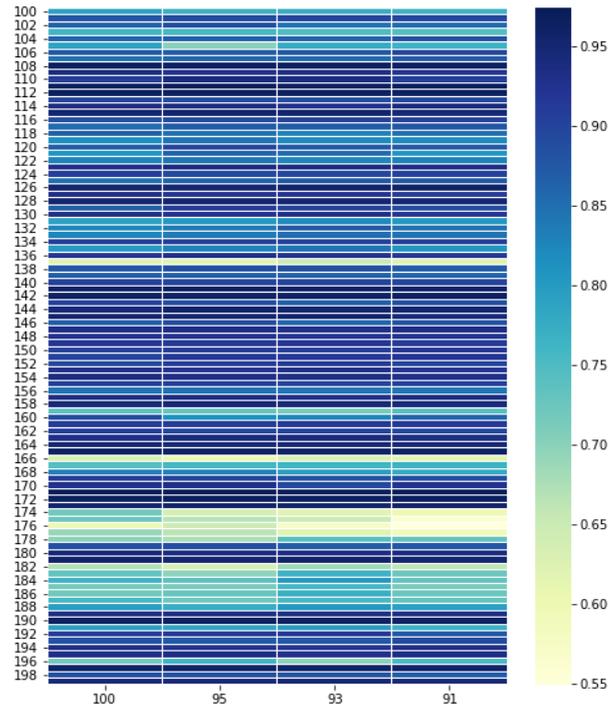
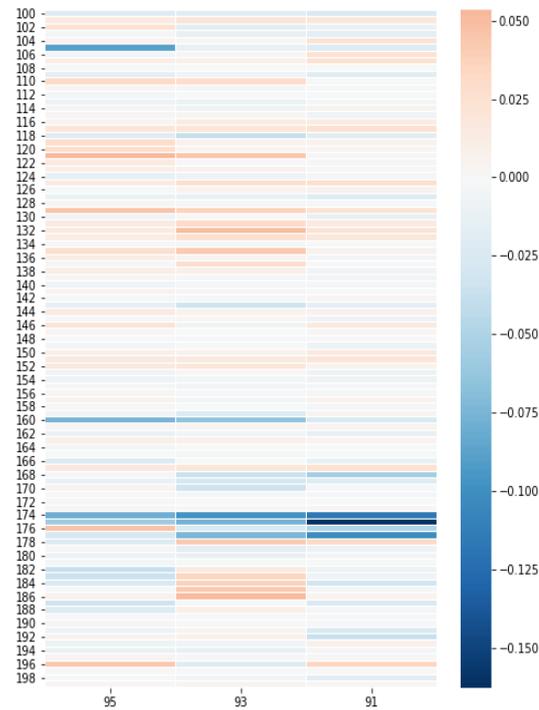


Fig. 3.5: Dendrogram of similarity matrix of 100 training cases was built by the average method. On the left side, we have cases with the best learning rate and on right the contrary results. Each row and column has a line, which unites it with similar results and forms a cluster



(a) Dice Score of test images prediction trained on full dataset (100) and reduced (95, 93 and 91)



(b) Subtraction between Dice Score of the original dataset and with reduced number of images: 95, 93 and 91 consequently, to illuminate cases where changes were the most significant. Red colour shows improvement, white remains invariable, blue — decline.

Fig. 3.6: Dice Score of test images prediction

4 Discussion

Nowadays it is no question that convolutional neural networks are state-of-art techniques for image recognition and segmentation. However, it remains some gaps for development and improving these models. From the test experiment, we may found several problems such as memory and GPU consumption. Despite huge breakthrough of Convolutional Neural Networks, it is still requiring an impressive amount of capacity. That was shown in the case where 3D U-Net did not perform well due to the necessity of the big GPU power machine. I may assume that many biomedical centres or hospitals cannot have access to such capacity in everyday life to train their model. It may be solved by pre-training or already prepared weights, which can be done remotely, but it will not allow the process to be more flexible and achieve the best segmentation results. Concerning the 2D U-Net we can easily conclude that it is a working example of CNN, which may be applied already in many machines and produce stable results, however they will be beaten by 3D version in future.

The results of experiment on 2D U-Net showed the use of input data adapted learning method leads to a slightly visible but in some cases a boost of segmentation performance. From several training case the median Dice Scores are shown an increase by 4.8% in case were 7 images were removed and near 1% of decrease in other cases. It is also visible that further improvement is possible if a better similarity classifier will be used, based on the fact that using a perfect similarity classifier (with already prepared ground truths) leads to 4% of improvement compared to the used method. Therefore, it let us conclude that there is a further potential for improving this technique.

We used a very basic approach for our challenge contribution, on the whole avoiding all post-processing. That was made in order to reduce the effect of the post-processing. Since the post-processing (brain extraction, noise reduction etc) seems to have a big influence in the final segmentation quality, a further increase could be achieved by an additional cleaning of the segmentation mask. From next ideas to improve process of optimazing the dataset could be work with meta-data of images, setups of scanners and others.

As the method does not depend on a given combination of features or a similarity classifier, it is possible to incorporate input data adapted learning method with most other, learning-based approaches. It is expected that most approach could benefit from the used method and results. Though we have to understand that developing a unique method of improving the segmentation of the image in convolutional neural networks quality is a generous task, which cannot be solved instantly and requires extensive discussion.

Conclusion

In the presented master's thesis was conducted a description of CT scans and MR images and the main differences between them on technical and image side. It was observed possible of deep learning methods and tools for image segmentation, explained the structure of convolutional neural networks on several examples and how they implemented into biomedical image analysis. There were provided a wide research of optimization and different loss functions for neural network. Further it includes a brief analysis of the training quality and method how to improve it.

The experiment part it about test training of neural network for brain tumour segmentation with state-of-art convolutional neural network nnU-Net on possible data loads and comparison of different loss functions. It shows us effectiveness and reliability in segmentation task, however, the quality can be improved. During the experiment, we trained 100 convolutional neural networks to assess the quality of predictions, and based on Dice Score calculations, built Similarity Matrices. From the training matrix, we constructed dendrogram (Fig.3.5) and clustered the images. Removing images from the worst cluster from the original dataset, we can achieve a similar median quality of the segmentation among all tested images, and in particular case, it was found an increase of almost 5 % as it showed in Tab.3.3. The discussion about training setups, an algorithm of clustering and similarity classifier, which unites data in a certain cluster, is behind the scope of one master's project, but the technique has a potential for development and will be in a future focus of scientific attention. The experiment has shown the success of the provided algorithm in detail analysis of performed training cases. The proposed outcome gives us certain ideas for future improving the quality of image segmentation via deep learning techniques.

Bibliography

- [1] Russakovsky O., Deng J., Su H., Krause J., Satheesh S., Ma S., Huang Z., Karpathy A., Khosla A., Bernstein M., Berg A. C., Fei-Fei L. *ImageNet Large Scale Visual Recognition Challenge* arXiv:1409.0575 (2014).
- [2] Osborn A., Digre K. *Imaging in neurology, 1st ed.*, p.4, Salt Lake City, UT ISBN 978-0-323-44781-2, (2016).
- [3] *Questions and Answers in MRI*, AD Elster, ELSTER LLC [accessed 5 Dec, 2019] Available from: <http://mriquestions.com/image-contrast-trte.html>.
- [4] Grueter B.E., Schulz U.G. *Age-related cerebral white matter disease (Leukoaraiosis)*, A review - Scientific Figure on ResearchGate, (2011). [accessed 10 Dec, 2019] Available from: https://www.researchgate.net/figure/CT-and-T2-weighted-MRI-showing-leukoaraiosis-at-different-stages-of-severity_fig1_51902532/.
- [5] Isensee F., Jäger P.F., Simon A. A. Kohl, Petersen J., Maier-Hein K.H. *nnU-Net: Breaking the Spell on Successful Medical Image Segmentation*, arXiv:1904.08128, (2019).
- [6] Hinkelmann K. *Neural Networks*, p. 7 University of Applied Sciences Northwestern Switzerland (2018) [accessed 15 Dec, 2019] Available from: http://didattica.cs.unicam.it/lib/exe/fetch.php?media=didattica:magistrale:kebi:ay_1718:ke-11_neural_networks.pdf.
- [7] Vinod N, Hinton G. E. *Rectified Linear Units Improve Restricted Boltzmann Machines*, 27th International Conference on International Conference on Machine Learning, ICML'10, USA: Omnipress, pp. 807–814, ISBN 9781605589077, (2010).
- [8] Eidnes L., Nøklund A. *Shifting Mean Activation Towards Zero with Bipolar Activation Functions* International Conference on Learning Representations (ICLR) Workshop. arXiv:1709.04054, Bibcode:2017 arXiv170904054E, (2018).
- [9] Maas A. L., Hannun A. Y., Ng A. Y. *Rectifier nonlinearities improve neural network acoustic models* Proc. ICML. 30 (1), (June 2013).
- [10] Xu B., Wang N., Chen T., Li M. *Empirical Evaluation of Rectified Activations in Convolutional Network*. arXiv:1505.00853, (2015-05-04).

- [11] Clevert D., Unterthiner T., Hochreiter S., *Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)*. arXiv:1511.07289, (2015-11-23).
- [12] Li W, Cao P, Zhao D, Wang J. *Pulmonary Nodule Classification with Deep Convolutional Neural Networks on Computed Tomography Images. Computational and mathematical methods in medicine.*, (2016). [accessed 10 Dec, 2019] Available from: <<https://www.hindawi.com/journals/cmmm/2016/6215085/>>.
- [13] Larrañaga P, Calvo B, Santana R, Bielza C, Galdiano J, Inza I, Lozano J, Armañanzas R, Santafé G, Pérez A, Robles V. *Machine learning in bioinformatics. Briefings in bioinformatics.* 7(1):86-112, (2006) . [accessed 10 Dec, 2019] Available from: <<https://www.ncbi.nlm.nih.gov/pubmed/16761367>>.
- [14] Garbade M. J. *Regression Versus Classification Machine Learning: What's the Difference?* (2018) [accessed 10 Dec, 2019] Available from: <<https://medium.com/quick-code/regression-versus-classification-machine-learning-whats-the-difference>>
- [15] Yusaku S. *Is the term "softmax" driving you nuts?*, Medium (2018) [accessed 10 Feb, 2020] Available from: <<https://medium.com/@u39kun/is-the-term-softmax-driving-you-nuts-ee232ab4f6bd>>
- [16] Ronneberger O., Fischer P., Brox T. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Computer Science Department and BIOS Centre for Biological Signalling Studies, University of Freiburg, Germany, arXiv:1505.04597v1, (2015).
- [17] Jaeger P.F., Kohl S. A. A., Bickelhaupt S., Isensee F., Kuder T. A., Schlemmer H.P., Maier-Hein K. H. *Retina U-Net: Embarrassingly Simple Exploitation of Segmentation Supervision for Medical Object Detection* arXiv:1811.08661, (21 Nov 2018).
- [18] He K., Gkioxari G., Dollar P., Girshick R. *Mask R-CNN* arXiv:1703.06870, (24 Jan 2018).
- [19] Murphy A. et al. *Supervised learning (machine learning)* (2019) [accessed 10 Dec, 2019] Available from: <<https://radiopaedia.org/articles/supervised-learning-machine-learning?lang=us>>
- [20] Bottou L., Bousquet O. *The Tradeoffs of Large Scale Learning*. In Sra, Suvrit; Nowozin, Sebastian; Wright, Stephen J. (eds.). *Optimization for Machine Learning*. Cambridge: MIT Press. pp. 351–368. ISBN 978-0-262-01646-9, (2012).

- [21] Qian, N. *On the momentum term in gradient descent learning algorithms*. Neural Networks : The Official Journal of the International Neural Network Society, 12(1), 145–151, (1999). [accessed 10 Feb, 2020] Available from: <[http://doi.org/10.1016/S0893-6080\(98\)00116-6](http://doi.org/10.1016/S0893-6080(98)00116-6)>
- [22] Nesterov, Y *A method for unconstrained convex minimization problem with the rate of convergence $o(1/k^2)$* . Doklady ANSSSR (translated as Soviet.Math.Docl.), vol. 269, pp. 543– 547, (1983).
- [23] Duchi, J., Hazan, E., Singer, Y. (2011). *Adaptive Subgradient Methods for Online Learning and Stochastic Optimization*. Journal of Machine Learning Research, 12, 2121–2159. [accessed 10 Feb, 2020] Available from: <<http://jmlr.org/papers/v12/duchi11a.html>>
- [24] Kingma D.P., Ba J. *Adam: A Method for Stochastic Optimization* arXiv:1412.6980, (30 Jan 2017).
- [25] Kingma, D. P., Ba, J. L. *Adam: a Method for Stochastic Optimization*. International Conference on Learning Representations, 1–13, (2015).
- [26] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S. *GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium*. In Advances in Neural Information Processing Systems 30, NIPS. (2017).
- [27] Ruder S. *An overview of gradient descent optimization algorithms* arXiv:1609.04747, (15 Jun 2017).
- [28] Isensee. F, Petersen J., Klein A., Zimmerer D., Jaeger P. F., Kohl S., Wasserthal J., Kohler G., Norajitra T., Wirkert S., and Maier-Hein K. H. *nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation*. Division of Medical Image Computing, German Cancer Research Center (DKFZ), Heidelberg, Germany arXiv:1809.10486, (2018).
- [29] Milletari, F., Navab, N., Ahmadi, S. *V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation*. Fourth International Conference on 3D Vision (3DV), Stanford, CA, 2016, pp. 565-571, (2016).
- [30] Salehi S. S. M., Erdogmus D., Gholipour A. *Tversky loss function for image segmentation using 3D fully convolutional deep networks*, (2017).
- [31] Berman M., Triki A. R., Blaschko M. B. *The Lovász-Softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks*, (2018).

- [32] Menze B. H., Jakab A., Bauer S., Kalpathy-Cramer J., Farahani K., Kirby J, et al. *The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)*, IEEE Transactions on Medical Imaging 34(10), 1993-2024, DOI: 10.1109/TMI.2014.2377694, (2015).
- [33] Isensee F., Kickingereder P., Wick W, Bendszus M. and Maier-Hein K. H. *No New-Net Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, Springer International Publishing, Cham, p. 234–244, ISBN 978-3-030-11726-9, (2019).
- [34] Ioffe S., Szegedy C. *Batch normalization: Accelerating deep network training by reducing internal covariate shift*. arXiv:1502.03167, (2015).
- [35] Goetz M., Weber C., Thiel C., Maier-Hein K. H. *Input Data Adaptive Learning (IDAL) for Sub-acute Ischemic Stroke Lesion Segmentation* Springer International Publishing Switzerland, A. Crimi et al. (Eds.): BrainLes 2015, p. 284-295. DOI: 10.1007/978-3-319-30858-6 25, (2016).
- [36] Simpson A.M., Antonelli M., Bakas S., Bilello M., Farahani K., Ginneken B., Kopp-Schneider A., Landman B. A., Litjens G., Menze B., Ronneberger O., Summers R. M., Bilic P., Christ P. F., Do R., Gollub M., Golia-Pernicka J., Heckers S., Jarnagin W., McHugo M., Napel S., Vorontsov E., Maier-Hein L., Jorge Cardoso M. *A large annotated medical image dataset for the development and evaluation of segmentation algorithms* arXiv:1902.09063, (25 Feb 2019).
- [37] Isensee F., Jaeger P., Wasserthal J., Zimmerer D., Petersen J., Kohl S., Schock J., Klein A., Ross T., Wirkert S., Neher P., Dinkelacker S., Koehler G, Maier-Hein K. *batchgenerators - a python framework for data augmentation*. DOI:10.5281/zenodo.3632567, (2020).
- [38] Tiu E. *Metrics to Evaluate your Semantic Segmentation Model*, Toward Data Science, (2019). [accessed 10 Feb, 2020] Available from: <<https://towardsdatascience.com/metrics-to-evaluate-your-semantic-segmentation-model-6bcb99639aa2>>
- [39] Sokal R.R., Michener C.D. *A statistical method for evaluating systematic relationships* University of Kansas Science Bulletin. 38: 1409–1438, (1958).

List of symbols, quantities and abbreviations

CNN	Convolutional Neural Network
CT	Computed Tomography
MR	Magnetic Resonance
fMRI	Functional Magnetic Resonance Image
SPECT	Single-photon Emission Computed Tomography
PET	Positron Emission Tomography
CSF	Cerebrospinal Fluid
FLAIR	Fluid-attenuation Inversion Recovery
ROI	Region of Interest
ReLU	Rectified Linear Unit