Automatic Brain Tumor Segmentation and Isocenter Selection

Zhehao Zhang[1*], Zhixu Tao[2], and Matthew Walker[3]

[1]Department of Statistics, University of Washington, Seattle, U.S.A.
[2]Department of Mathematics, University of Toronto, Toronto, Canada.
[3]Neuroscience PhD, University of Toronto, Toronto, Canada.
[*]Corresponding author. Email: zhehaoz@uw.edu
[2]zhixu.tao@mail.utoronto.ca
[3]matthew.walker21@gmail.com
[†]These authors contributed equally to this work.

Abstract

Brain Magnetic Resonance Imaging (MRI) are widely used in clinics to perform brain tumor diagnosis. For radiotherapy, experienced physicians need to segment the tumor(s) on MRI images first and then decide the isocenters for radiation. However, the manual segmentation process is both time-consuming and requires years of experience. Thus, modern computer-vision based techniques can be particularly helpful in this process. This paper provides a comprehensive examination of current automatic segmentation techniques on brain MRI images. We also discuss automatic isocenter selection process based on the segmentation of tumor. These should serve as an overview for future research and development in related topics.

1 Introduction

Brain related tumors are commonly treated with high-dose radiation beams from Gamma Knife. These treatments are delivered over a single or a few sessions and are referred to as stereotactic radiosurgery (SRS) treatment plans [1]. During Gamma Knife treatment, beams from certain radioactive sources are focused so that they intersect at the same location in space and the result will be a spherical region of high dose which is termed as a shot of radiation [2]. There are four different shot sizes. The center at which those beams intersect is called isocenter. In order to deliver efficient SRS treatment, the image segmentation of brain tumors and very precise treatment plan are required to ensure that radiation beams can cover the target area, namely, the tumor as closely as possible. This paper discusses several methods that are commonly used to automatically segment brain images and select isocenters for treatment plan.

Before deep learning took over computer vision, popular approaches for automatic brain tumor

segmentation involves Random Forest, Support Vector Machine (SVM), Region Growing, k-Nearest Neighbor (KNN). Wadhwa (2019) [3] "A review on brain tumor segmentation of MRI images" Section II summarizes some representatives of these models in brain MRI segmentation application.

In recent years, deep learning architecture and especially convolutional neural networks (CNN) outperforms the traditional models and gains popularity in this task. In particular, Fully Convolutional Networks (Long et al., 2015 [4]) has become the state of art for this job due to its computational efficiency and the ability to produce segmentation maps.

Different from usual image classification, medical image segmentation requires dense pixel-wise predictions with delineation of boundary. Pooling layers in CNN that contract local information and improve computational efficiency becomes a major problem in our task. One of the popular solution for this is an encoder-decoder structure with prior information stack to later layers such as U-net (Ronneberger et al., 2015[5]). U-net is developed specifically for medical image segmentation tasks and can process both 2D images and 3D volumes (also known as V-net). Another popular method in brain segmentation field is by Kamnitsas (2017) [6] with a muti-scale 3D CNN with a fully connected Conditional Random Field (CRF). The method wins the BraTS competition that year and was later developed into a software called DeepMedic. However, even though new architectures keep evolving in recent years, the computational complexity for this task is still a problem. Most convolutional neural network models take days or even weeks to train on some of the most advanced GPUs nowadays. Noticeably, Wacker (2019) [7] applied transfer learning to this task and achieved some success. The pretrained networks is adapted from ResNet34 on ImageNet dataset and decoding blocks are added. The result is competitive among the most popular methods.

Before methods of selecting isocenters, there were many different methods to optimize the Gamma Knife treatment plan. For example, Wang 1999 [8] regarded automated radiosurgical treatment planning as an optimization problem of packing unequal spheres into a three-dimensional bounded region. Given a 3D bounded region, a positive integer V, a multiset of spheres and a location constraint on spheres, Wang proposed a method to find the minimum number of spheres such that the covered volume is at least V and the location constraint is satisfied. Wang's method is based on sphere-packing algorithm and later, there are more and more papers discussing automatic treatment plan based on sphere-packing algorithm such as Wagner et al. 2000 [9] and Ghobadi et al. 2012 [10]. In 2000, Ferris et al. [2] proposed a dose model which formulated treatment planning process as an optimization model. In this optimization model, the basic variables include the number of shots, the width of the shot, the coordinates of the center location of the shot and the duration of each shot of radiation. Also, two approaches were proposed to solve this optimization problem, namely using mixed integer programming (MIP) and non-linear programming (NIP).

Different than traditional sphere packing algorithms and optimization models which consider several different variables, in this paper, we discuss methods that only consider the location of isocenters. Moreover, as deep learning becomes more and more popular, a method that uses neural networks

for predicting treatment plans will also be introduced.

2 Brain Tumor Segmentation

2.1 Datasets

To train a deep neural network model to segment a type of tumor, usually we need at least $100-200$ sample MRI scans. For brain tumor, the two most popular public datasets include BraTS dataset [11] and Human Connectome Project (HCP) dataset. BraTS is also a competition held by University of Pennsylvania on tumor segmentation. The dataset contains more than 200 objects each with co-registered T1, T1ce, T2 and FLAIR images and manual segmentation as ground truth. The tumors are also classified as High Grade Glioma (HGG) and Low Grade Glioma(LGG). HCP dataset contains structural T1 and T2 scans for more than 1000 objects. Other datasets are more local and might be tumor-specific. For example, the dataset at Toronto Western Hospital focuses specifically on acoustic neuroma with sensitive region on cochlea. The Organ at Risk (OAR) delineation for datasets may differ due to different brain tumor types.

2.2 U-Net

One of the most popular architecture for brain tumor segmentation is the encoder-decoder structure. We took a brief look at BraTS competition winners and we can see in 2018 and 2019, the first place in both years and the second place in 2018 all used this kind of architecture. One of the main advantage of encoder-decoder structure is the computational efficiency. The pooling layer in CNN still works and we can avoid a very deep convolutional neural network without pooling. The second place in 2018 BraTS dataset even published a paper called "No New Net" (Isensee et al., 2018 [12]) arguing that a well trained U-Net is hard to beat in this task with high efficiency. We adopted a relatively simple 2D U-Net structure by Dong et al. (2017) [13]. The model takes a slice-by-slice approach and is relatively cheap computationally. Figure 1. shows the general architecture of the model.

We used the Compute Canada parallel computing with Graham server located at Waterloo, Canada. A whole node of T4 GPU is used to train the model. BraTS 2019 dataset was used as training set in this case. The following shows some of the selected testing results based on BraTS 2020 objects. "iou" stands for intersection over union for segmented tumor and ground truth, which is a relatively intuitive measure for how good our segmentation is.

3 Cross-Modality Image Translation

A lot of existing models are based on multi-modality scans like T1, T2, Flair etc. Each modality presents a unique view or function of images. From Yang et al. (2020) [14], T1 images are favorable

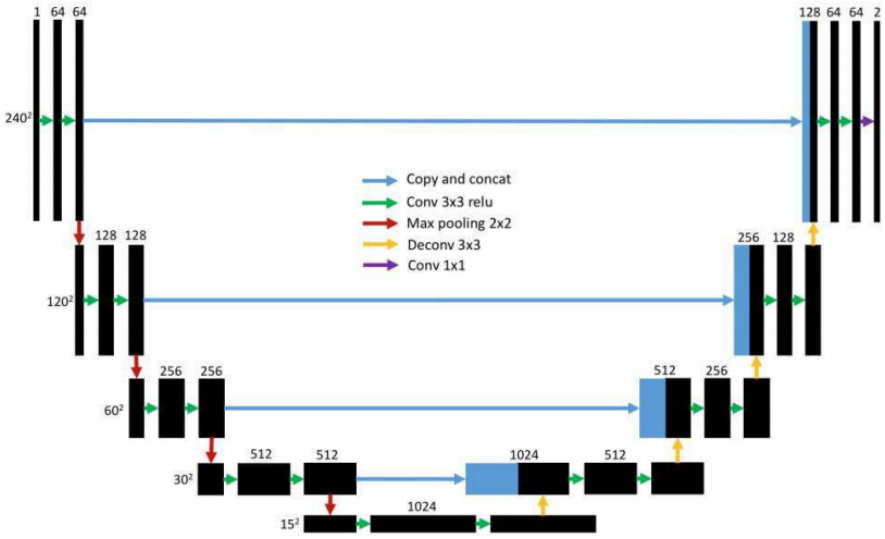

Figure 1: Dong's U-net Architecture [13]

for observing structures, e.g. gray matter and white matter in the brain; T2 images are utilized for locating tumors; T2-Flair images present the location of lesions with water suppression. However, the acquisition of scans may be different for each organization and some are more difficult than others. Modern deep learning and especially the Generative Adversarial Networks (GAN) provides a new possibility for this problem. From 2018 and on, architectures like conditional GAN have been applied for image modality translation in brain MRI field. To compare the generated image with ground truth, a measure for similarities between two distributions is needed. The most popular approaches are KL (Kullback–Leibler) divergence divergence and Wasserstein distance. The KL divergence for probability distributions p, q are:

$$D_{KL}(p||q) = \int_x p(x) \log \frac{p(x)}{q(x)} dx$$

which achieves minimum 0 when p and q are identical almost everywhere. The Wasserstein distance has a more advanced definition and it works better than KL divergence for most of times. Wasserstein GAN (WGAN) is specifically designed to apply Wasserstein distance to a GAN. I would recommend John-Niles Weed's video for a detailed introduction to Wasserstein metric. Maybe in the near future, we do not need to take multiple scans of MRI and one is enough to generate others.

Han et al. (2018) [15] uses WGAN to generate synthetic Brain MR image and passes the visual Turing test (a physician cannot tell which one is synthetic). The generated images can be used for data augmentation, physician training or treatment improvement. Conditional GAN approach is used for image modality translation. Welander et al. (2018) [16] compares CycleGAN and UNIT for this task and conclude that a modified version of CycleGAN is very good at corss-modality image translation. A major work in this field is by Yang et al. (2018) [14] on Nature Scientific Report. They use the following conditional GAN architecture for modality translation: The code is modified

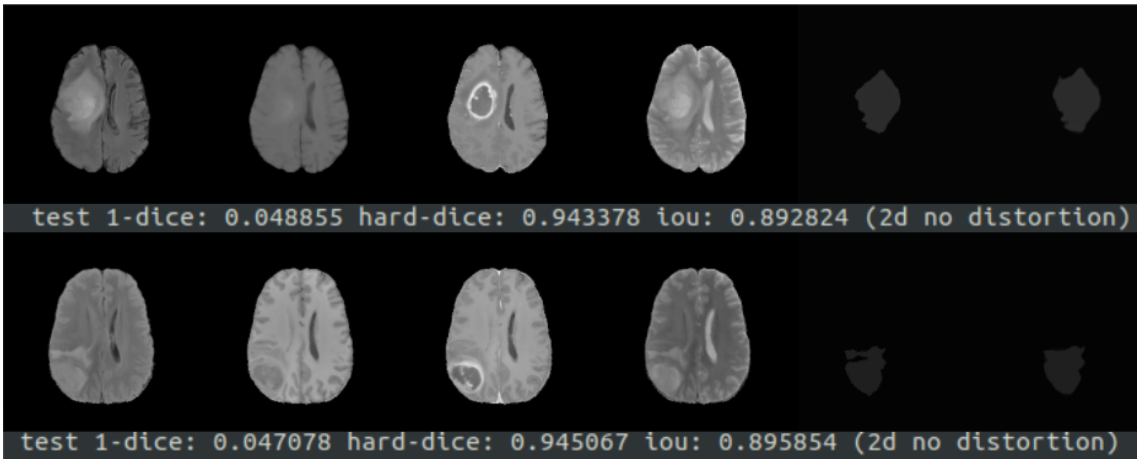

Figure 2: Selected segmentation testing result on BraTS 2020 dataset by Dong's U-net architecture.

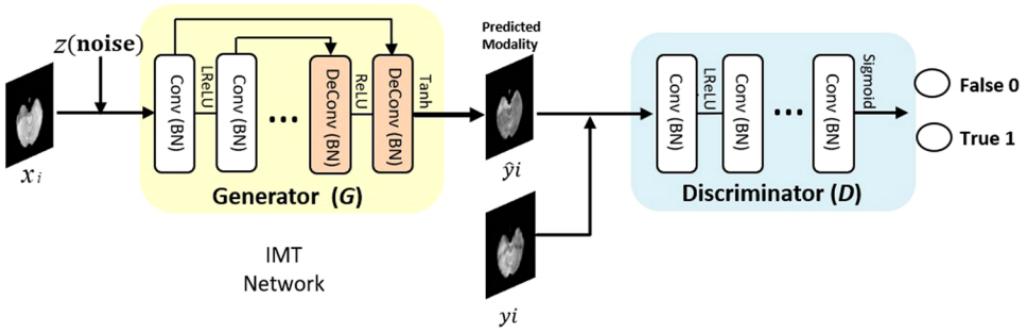

Figure 3: Cross-Modality Image Translation by Yang et al.

from https://github.com/phillipi/pix2pix and they compare the performance with different public datasets. The experimental results show promising applicability in real life and it can be used to translate missing scans or in fact any scan we need.

4 Isocenter Selection

In Gamma Knife Radiosurgery, high intensity gamma radiation will be used to kill brain tumors. The brain tumor is the target area of a treatment plan. In order to produce an optimal treatment plan, we need to make sure the following:

- Shots can cover the target area as much as possible.
- Shots cannot hurt organs at risk.
- Shots cannot overlap with each other.
- The number of shots should be as small as possible.

Therefore, variables in this optimization problem may include the number of shots, duration of radiation, the location of shots and the size of shots etc. In this section, we will only focus on the position of shots, i.e., the location of isocenters. After isoceneter selection, these isocenters will be used in the sector duration optimization (SDO) model. This section will discuss about several different methods that have been used so far to optimize the isocenter selection.

4.1 Grassfire and Sphere Packing Algorithm

The sphere packing problem tries to pack spheres into a given geometric domain. Since each shot is a spherical region, the optimization problem can be formulated as a sphere packing problem. Several literature (Wagner et al. 2000 [9]; Ghobadi et al. 2012 [10]) discussed about a hybrid grassfire and sphere-packing algorithm (GSP) that is used to find isocenters.

The GSP algorithm that we will discuss in this section is proposed by Ghobadi et al. in 2012 [10]. In image processing, the grassfire algorithm (Blum 1967 [17]) is used to compute the distance of a pixel to the border of the given region. The grassfire algorithm helps to extract an object's skeleton. A skeleton is a representation of an object that contains all information needed to reconstruct the original object. The aim of using grassfire algorithm here is to find the deepest voxels in the current target volume. The deepest-position voxels may not be unique. The grassfire algorithm may select several voxels as candidates. It then uses a scoring metric to select the best one voxel with higher score as an isocenter. If those candidates have the same score, one is chosen randomly. At this newly selected isocenter, a shot, i.e., a sphere of the largest possible size (among those four sizes) will fit inside the target. Next, voxels covered by the shot centered at the isocenter will be removed. This step will guarantee that there is no overlap between shots. This algorithm will run iteratively until the desired number of isocenters is achieved or the percentage of uncovered voxels falls below an acceptable number.

The scoring method we use here to select the isocenter among the set of deepest voxels is based on two factors:

- a score related to the distance between voxel i and the set of previously selected isocenters,
- a score related to the irregularity of the target and the distance between voxel i and the boundary of the target.

Figure 4 shows an example of isocenter selection on a representative clinical target structure by running GSP algorithm. In figure 4(a), grassfire algorithm identifies the set of deepest voxels and the scoring metric helps select the best isocenter. In figure 4(b), a suitable sphere fits into the target volume and those voxels covered by this sphere are removed. In figure 4(c), grassfire algorithm applied again to the new target volume and a new isocenter is identified.

This algorithm helps produce the optimal treatment plan significantly, but it also has disadvantages. This algorithm cannot address the challenge caused by symmetry of a target very well. It

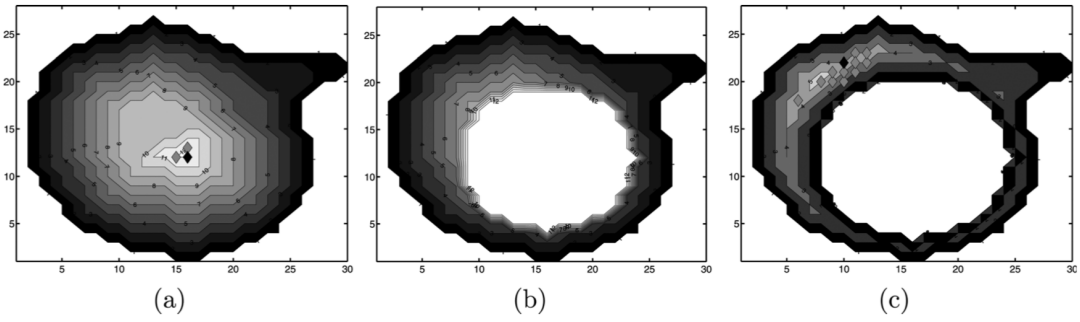

Figure 4: Isocenter selection on a representative clinical target structure [10].

is proven that for a symmetric shape, sphere-packing algorithm can be far from the optimal treatment plan (Wang 2000 [18]). For example, the skeleton of a sphere is just the center of the sphere. However, irradiating only one point is usually not desirable since the dose cannot cover a sufficient fraction of the target volume. In order to address this challenge, we introduce an improved version of skeletonization algorithm (Doudareva et al. 2015 [19]).

4.2 Improved Skeletonization for Isocenter Selection

Traditional skeletonization technique that we discussed in section 3.1 only extracts the skeleton of an object, so it cannot deal with symmetric tumors very well. In 2015, Doudarerva et al. [19] designed an improved version of skeletonization algorithm which solves the problem imposed by symmetry of a tumor. Instead of extracting an object's skeleton, they define skeletal clusters based on voxel depths by using grassfire algorithm and identify multiple clusters within the target volume. This improved algorithm first identify the deepest voxels inside the current target volume and those voxels will be considered as a skeletal cluster. Once the skeleton is identified, it will be subsequently removed from the target. However, every cluster is only removed for k iterations and here k is provided as user input. In other words, those removed skeletons will be added back to the target volume after k iterations. Compared with the algorithm that constantly removes voxels without adding them back, this new algorithm uses data re-insertion to help us better capture the shape of tumor by reusing previously deleted voxels. Therefore, this algorithm can ensure that selected isocenters will be distributed over the whole target volume.

Figure 5 [19] shows the first four iterations of this skeletonization algorithm when k is 2.In these graphs, darkness represents the depth of target voxels. Figure 5(a) shows the first iteration where the first skeleton is identified and labeled as 1. In iteration 2 (figure 5(b)), the skeleton 1 is removed from the current target volume and since k is 2, it will not be added back until the fourth iteration. Also, in this iteration, the second skeleton, skeleton 2, is identified. In iteration 3 (figure 5(c)), skeleton 2 is removed from the current target volume for the next two iterations and will be added back at the fifth iteration which is not shown in this figure. At the same time, skeleton 3 is identified. In iteration 4 (figure 5(d)), skeleton 1 is added back. The algorithm will continue in this manner

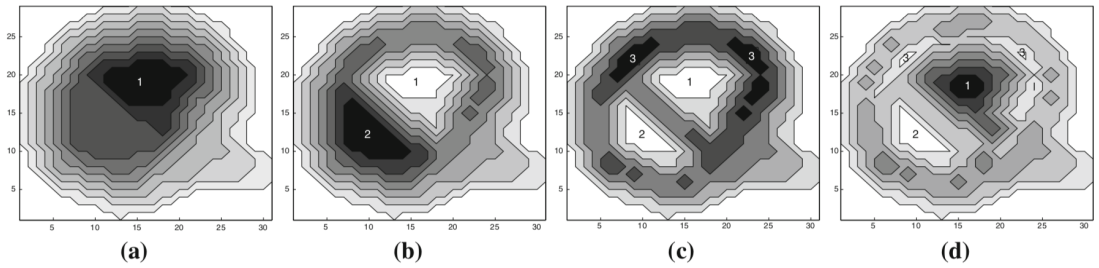

Figure 5: the first four iterations of the skeletonization algorithm

until the desired number of isocenters is reached.

We also notice that this data re-insertion step may lead to overlapping of clusters. This may result in isocenters that are close to each other. However, this problem will be solved later in the SDO model. Figure 6 is a comparison of isocenter selection for a spherically shape target by using two different algorithms. Figure 6(a) is given by conventional skeletonization algorithm while figure 6(b) is given by this new skeletonization algorithm. It is obvious that in figure 6(b), the use of multi-clusters produce better distribution of isocenters.

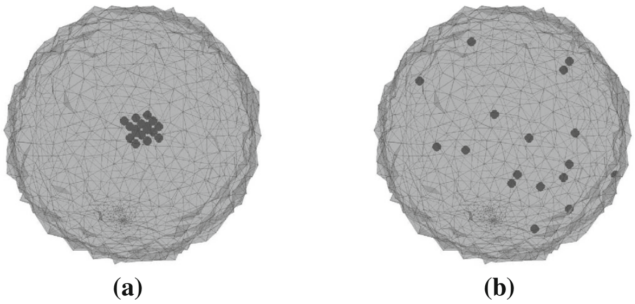

Figure 6: comparison of two algorithms

4.3 Knowledge-based Isocenter Selection

Recently, techniques from machine learning and deep learning are widely used in image processing and radiosurgery. In 2020, Berdyshev et al. [20] proposed a knowledge-based isocenter selection algorithm by using residual neural network.

The residual neural network is used to alleviate the problem caused by training very deep neural networks. Very deep neural networks will have vanishing gradient problem, namely, as the gradient is back propagated to the previous layers, repeated multiplications may make the gradient infinitely small. This will influence the performance of the deep neural networks. Therefore, when there are

too many layers in the neural network, the value of cost function will begin increasing compared with neural networks with less layers. With residual neural networks, it is possible to train hundreds or even thousands of layers. The key idea of the residual net is so called the identity shortcut connection. This identity shortcut connection allows the neural net to take the activation function from one layer and suddenly feed it to another layer. Figure 7 is an example of short identity connection.

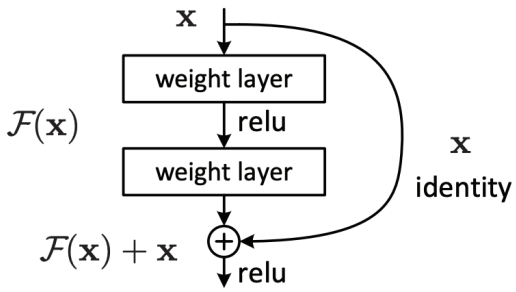

Figure 7: An example of short identity connection [21]

For a segmented image, our target area may not be connected in this image. Therefore, in data preprocessing step, we will first partition the target area into several separated subsets so that each subset is connected in the segmented image. Then a collection of tumor spaces will be constructed based on these separated subsets. These tumor spaces are used to compute shape descriptors. A shape descriptor is a vector that can be used to recover the image. With these shape descriptors, the input of the residual neural nets will no longer be volumetric images. Instead, the input will be a set of shape descriptors for tumor spaces and isocenters. With the help of shape descriptors, the training of neural nets will be much more efficient.

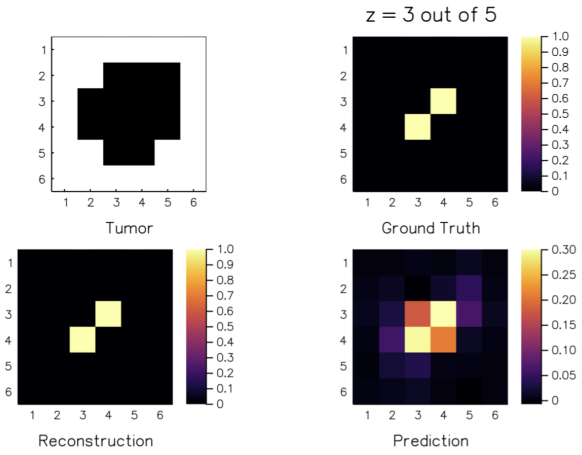

Figure 8: An example of prediction [20]

Figure 8 including four heat maps shows an example of prediction from test cases. In figure 8, Tumor

is a slice of tumor, ground truth is the manually selected isocenter for the tumor slice, reconstruction is the reconstructed image from true shape descriptors for the selected isocenters, and prediction is the reconstructed image from predicted isocenter shape descriptors. In these heat maps, the relative scale is much more meaningful compared with the absolute scale.

5 Discussion

While U-Net gains its popularity from its simple architecture and efficient segmentation performance, there are deeper reasons why it works so well on medical image. https://github.com/ShawnBIT/UNet-family gives a summary of architectures in U-Net family and briefly explains why it works. Like most segmentation architectures, U-Net can also been seen as a revised version of FCN. The main difference is that it introduces concatenation and encoder-decoder structure. For medical images, usually we have very small data sets. 100 objects is very typical for a training set. In this case, a complicated architecture is quite easy to over-fit. Also, medical image contains lower variation than usual image classification datasets. The brain scan only contains CT or MRI of the brain rather than other areas. With some pre-registration, the variability of medical images is relatively low. We can also use data augmentation techniques, patches by patches segmentation to increase our sampling size.

Apart from segmenting tumors from MRI scans or CT scans, sometimes we also want to segment any Organs at Risk (OAR) to avoid important organs being affected by radiation during therapy. Agn et al. (2019)[22] proposed a method by using convolutional restricted Boltzmann Machine to segment tumor and OAR at the same time. van der Heyden et al. (2019) [23] also studies the segmentation of OAR on dual-energy CT. In general, this might be a more complicated classification task if we want to combine the segmentation of OAR with tumor and isocenter selection because the segmentation would be related to location of the tumor and treatment. A multi-class classification is also required in this task.

For optimizing radiosurgery treatment planning, another method is to solve fully connected quadratic unconstrained binary optimization (QUBO) by Fujitsu Digital Annealer (DA) [24]. Digital Annealer is a new architecture that solves combinatorial optimization problems at high speed with digital circuits inspired by quantum phenomena. The Digital Annealer's algorithm is based on the annealing metal processing phenomenon. When exploring optimal solutions, annealing method first search all solutions including those far from optimal and then gradually close in to an optimal solution. Currently it can solve problems of up to 1,024 variables. For the improved skeletonization algorithm for isocenter selection that we discussed in section 4.2, how to transform the algorithm into the digital annealing's algorithm is still unsolved.

6 Resources

The following links are some resources that might be helpful:
A documentation about segmentation and deep learning related works: Link

A guide on how to use the Compute Canada computing resources and supercomputers: Link
Code for training and testing BraTs 2019 Dataset on Compute Canada: Link

Acknowledgments

We would like to acknowledge and thank our supervisor, Mojgan Hodaie, as well as all group members, our friends and families. We would also like to thank the ComputeCanada system, the dcm2nixx package, Atom editor and Python3.

References

[1] K. Ghobadi, "Optimization methods for patient positioning in leksell gamma knife perfexion," PhD thesis, 2014.

[2] M. C. Ferris and D. M. Shepard, "Optimization of gamma knife radiosurgery," 2000.

[3] A. Wadhwa, A. Bhardwaj, and V. S. Verma, "A review on brain tumor segmentation of mri images," *Magnetic resonance imaging*, vol. 61, pp. 247–259, 2019.

[4] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241.

[6] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation," *Medical image analysis*, vol. 36, pp. 61–78, 2017.

[7] J. Wacker, M. Ladeira, and J. E. V. Nascimento, "Transfer learning for brain tumor segmentation," *arXiv preprint arXiv:1912.12452*, 2019.

[8] J. Wang, "Packing of unequal spheres and automated radiosurgical treatment planning," *Journal of Combinatorial Optimization*, vol. 3, no. 4, pp. 453–463, 1999.

[9] T. H. Wagner, T. Yi, S. L. Meeks, F. J. Bova, B. L. Brechner, Y. Chen, J. M. Buatti, W. A. Friedman, K. D. Foote, and L. G. Bouchet, "A geometrically based method for automated radiosurgery planning," *International Journal of Radiation Oncology* Biology* Physics*, vol. 48, no. 5, pp. 1599–1611, 2000.

[10] K. Ghobadi, H. R. Ghaffari, D. M. Aleman, D. A. Jaffray, and M. Ruschin, "Automated treatment planning for a dedicated multi-source intracranial radiosurgery treatment unit using projected gradient and grassfire algorithms," *Medical physics*, vol. 39, no. 6Part1, pp. 3134–3141, 2012.

[11] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, R. T. Shinohara, C. Berger, S. M. Ha, M. Rozycki, *et al.*, "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge," *arXiv preprint arXiv:1811.02629*, 2018.

[12] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. Maier-Hein, "No new-net," in *International MICCAI Brainlesion Workshop*, Springer, 2018, pp. 234–244.

[13] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, "Automatic brain tumor detection and segmentation using u-net based fully convolutional networks," in *annual conference on medical image understanding and analysis*, Springer, 2017, pp. 506–517.

[14] Q. Yang, N. Li, Z. Zhao, X. Fan, E. I. Chang, Y. Xu, *et al.*, "Mri cross-modality neuroimage-to-neuroimage translation," *arXiv preprint arXiv:1801.06940*, 2018.

[15] C. Han, H. Hayashi, L. Rundo, R. Araki, W. Shimoda, S. Muramatsu, Y. Furukawa, G. Mauri, and H. Nakayama, "Gan-based synthetic brain mr image generation," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, IEEE, 2018, pp. 734–738.

[16] P. Welander, S. Karlsson, and A. Eklund, "Generative adversarial networks for image-to-image translation on multi-contrast mr images-a comparison of cyclegan and unit," *arXiv preprint arXiv:1806.07777*, 2018.

[17] H. Blum *et al.*, *A transformation for extracting new descriptors of shape*. MIT press Cambridge, 1967, vol. 4.

[18] J. Wang, "Medial axis and optimal locations for min-max sphere packing," *Journal of combinatorial optimization*, vol. 4, no. 4, pp. 487–503, 2000.

[19] E. Doudareva, K. Ghobadi, D. M. Aleman, M. Ruschin, and D. A. Jaffray, "Skeletonization for isocentre selection in gamma knife® perfexion™," *Top*, vol. 23, no. 2, pp. 369–385, 2015.

[20] A. Berdyshev, M. Cevik, D. Aleman, H. Nordstrom, S. Riad, Y. Lee, A. Sahgal, and M. Ruschin, "Knowledge-based isocenter selection in radiosurgery planning," *Medical Physics*, 2020.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[22] M. Agn, P. M. af Rosenschöld, O. Puonti, M. J. Lundemann, L. Mancini, A. Papadaki, S. Thust, J. Ashburner, I. Law, and K. Van Leemput, "A modality-adaptive method for segmenting brain tumors and organs-at-risk in radiation therapy planning," *Medical image analysis*, vol. 54, pp. 220–237, 2019.

[23] B. van der Heyden, P. Wohlfahrt, D. B. Eekers, C. Richter, K. Terhaag, E. G. Troost, and F. Verhaegen, "Dual-energy ct for automatic organs-at-risk segmentation in brain-tumor patients using a multi-atlas and deep-learning approach," *Scientific reports*, vol. 9, no. 1, pp. 1–9, 2019.

[24] M. Aramon, G. Rosenberg, E. Valiante, T. Miyazawa, H. Tamura, and H. G. Katzgraber, "Physics-inspired optimization for quadratic unconstrained problems using a digital annealer," *Frontiers in Physics*, vol. 7, p. 48, 2019.