



Tumor Segmentation and Survival Prediction in Glioma with Deep Learning

Li Sun¹, Songtao Zhang¹, and Lin Luo²(✉)

¹ Southern University of Science and Technology, Shenzhen 518055, China

² Peking University, Beijing 100871, China
luo1@pku.edu.cn

Abstract. Every year, about 238,000 patients are diagnosed with brain tumor in the world. Accurate and robust tumor segmentation and prediction of patients' overall survival are important for diagnosis, treatment planning and risk factor characterization. Here we present a deep learning-based framework for brain tumor segmentation and survival prediction in glioma using multimodal MRI scans. For tumor segmentation, we use ensembles of three different 3D CNN architectures for robust performance through majority rule. This approach can effectively reduce model bias and boost performance. For survival prediction, we extract 4524 radiomic features from segmented tumor region. Then decision tree and cross validation are used to select potent features. Finally, a random forest model is trained to predict the overall survival of patients. On 2018 MICCAI Multimodal Brain Tumor Segmentation Challenge (BraTS), our method ranks at second place and 5th place out of 60+ participating teams on survival prediction task and segmentation task respectively, achieving a promising 61.0% accuracy on classification of long-survivors, mid-survivors and short-survivors.

Keywords: Survival prediction · Brain tumor segmentation · 3D CNN · Multimodal MRI

1 Introduction

Brain tumor is cancerous or noncancerous mass or growth of abnormal cells in the brain, malignant brain tumor is one of the most aggressive and fatal tumors. Originated in the glial cells, gliomas are the most common brain tumor. [6] Depending on the pathologic evaluation of the tumor, gliomas can be categorized into glioblastoma (GBM/HGG) and lower grade glioma (LGG). Gliomas contain various heterogeneous histological sub-regions, including peritumoral edema, necrotic core, enhancing and non-enhancing tumor core. Magnetic resonance imaging (MRI) is commonly used in radiology to portray the phenotype and intrinsic heterogeneity of gliomas, since multimodal MRI scans, such as T1-weighted, contrast enhanced T1-weighted (T1c), T2-weighted and Fluid Attenuation Inversion Recovery (FLAIR) images, provide complementary profiles for

different sub-regions of gliomas. For example, the enhancing tumor sub-region is described by areas that show hyper-intensity in T1Gd scan when compared to T1 scan.

Accurate and robust prediction of overall survival through automated algorithms for patients diagnosed with gliomas can provide valuable guidance for diagnosis, treatment planning and outcome prediction. However, the selection of reliable and potent prognostic is difficult. Medical imaging (e.g. MRI, CT) can provide radiographic phenotype of tumor, and it has been exploited increasingly to extract and analyze quantitative imaging features. [7] Clinical data, including patient age, resection status and others, also provide important information about patients' outcome.

Segmentation of gliomas in pre-operative MRI scans, conventionally done by expert board-certified neuroradiologists, can provide quantitative morphological characterization and measurement of gliomas sub-regions. It is also pre-requisite for survival prediction since most potent features are derived from the tumor region. This quantitative analysis has great potential for diagnosis and research, as it can be used for grade assessment of gliomas and planning of treatment strategies. But this task is challenging due to the high variance in appearance and shape, ambiguous boundaries and imaging artifacts. Until now, automatic segmentation of brain tumors in multimodal MRI scans is still one of the most difficult tasks in medical image analysis. In recent years, deep convolutional neural networks (CNNs) have achieved great success in the field of computer vision. Inspired by the biological structure of visual cortex, CNNs are artificial neural networks with multiple hidden convolutional layers between the input and output layers. They have non-linear property and are capable of extracting higher level representative features. CNNs have been applied into a wide range of fields and achieved state-of-the-art performance on tasks such as image recognition, instance detection, and semantic segmentation.

In this paper, we present a novel deep learning based framework to segment brain tumor and its subregion from MRI scans, then perform survival prediction based on radiomic features extracted from segmented tumor sub-regions as well as clinical feature. Our automatic framework for brain tumor segmentation and survival prediction ranks at second place and 5th place out of 60+ participating teams on survival prediction task and segmentation task on 2018 MICCAI BraTS Challenge respectively, achieving a promising 61.0% accuracy on classification of long-survivors, mid-survivors and short-survivors.

2 Methodology

2.1 Overview

Our proposed framework for survival prediction using MRI scans consists of the following steps, as illustrated in the figure below. First, tumor subregions are segmented using an ensemble model comprising of three different convolutional neural network architectures for robust performance through voting/majority rule. Then radiomics features are extracted from tumor sub-regions and total

tumor volume. Next, decision tree regressor with gradient boosting is used to fit the training data and rank the importance of each feature based on variance reduction, and cross validation is used to choose the optimal number of top-ranking features to use. Finally, a random forest model is used to fit the training data and predict the overall survival of patient (Fig. 1).

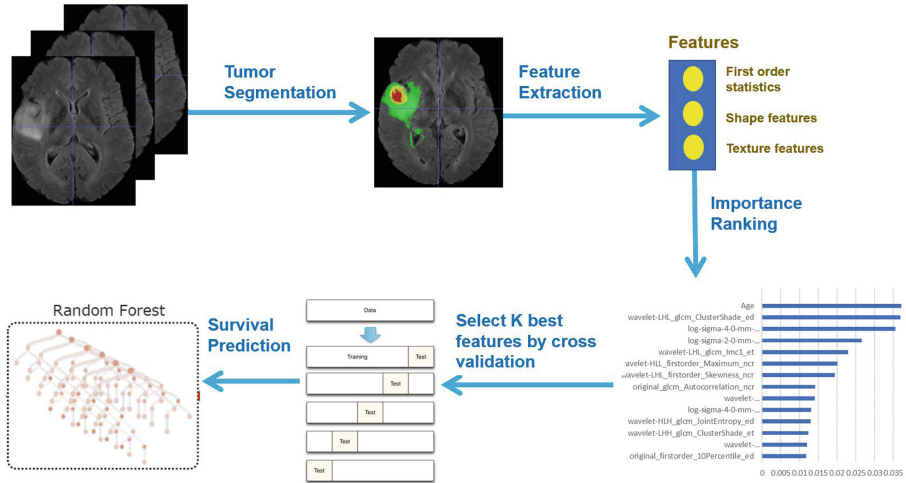


Fig. 1. Framework overview

2.2 Data Preprocessing

Since the intensity value of MRI is dependent on the imaging protocol and scanner used, we applied intensity normalization to reduce the bias in imaging. More specifically, the intensity value of each MRI is subtracted the mean and divided by the standard deviation of the brain region. In order to reduce overfitting, we applied random flipping and random gaussian noise to augment the training set.

2.3 Network Architecture

In order to perform accurate and robust brain tumor segmentation, we use an ensemble model comprising of three different convolutional neural network architectures. A variety of models have been proposed for tumor segmentation. Generally, they differ in model depth, filter number, connection way and others. Different model architectures can lead to different model performance and behavior. By training different kinds of model separately and merge the result, the model variance can be decreased and the overall performance can be improved. [11] We use three different CNN models and fuse the result by voting/majority rule. The detailed description of each model will be discussed as follows.

CA-CNN. The first network we employ is Cascaded Anisotropic Convolutional Neural Network (CA-CNN) proposed by Wang et al. [17]. The cascade is used to convert multi-class segmentation problem into a sequence of three hierarchical binary segmentation problems. The network is illustrated as follows (Fig. 2):

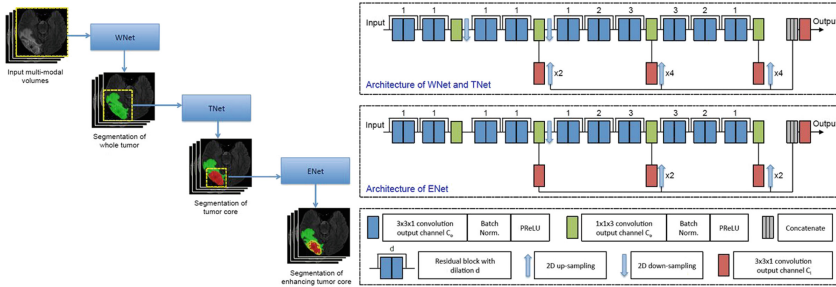


Fig. 2. Cascaded framework and architecture of CA-CNN

This architecture also employs anisotropic and dilated convolution filters, which are combined with multi-view fusion to reduce false positives. It also employs residual connections [8], batch normalization [9] and multi-scale prediction to boost the performance of segmentation. For implementation, we train the CA-CNN model using Adam optimizer, and set Dice coefficient as loss function. We set initial learning rate to 1×10^{-3} , weight decay 1×10^{-7} , batch size 5, and maximal iteration $30k$.

DFKZ Net. The second network we employ is DFKZ Net, which was proposed by Isensee et al. [10] from German Cancer Research Center (DFKZ). This network is inspired by U-Net. It employs a context encoding pathway that extracts increasingly abstract representations of the input, and a decoding pathway used to recombine these representations with shallower features to precisely segment the structure of interest. The context encoding pathway consists of three content modules, each has two $3 \times 3 \times 3$ convolutional layers and a dropout layer with residual connection. The decoding pathway consists of three localization modules, each contains a $3 \times 3 \times 3$ convolutional layer followed by a $1 \times 1 \times 1$ convolutional layer. For the decoding pathway, the output of layers of different depth is integrated by elementwise summation, thus the supervision can be injected deep in the network (Fig. 3).

For implementation, we train the network using Adam optimizer. To address the problem of class imbalance, we utilize the multi-class Dice loss function [10]:

$$L = -\frac{2}{|K|} \sum_{k \in K} \frac{\sum_i u_{i(k)} v_{i(k)}}{\sum_i u_{i(k)} + \sum_i v_{i(k)}} \quad (1)$$

where u denotes output possibility, v denotes one-hot encoding of ground truth, k denotes the class, K denotes the total number of classes and $i(k)$ denotes the

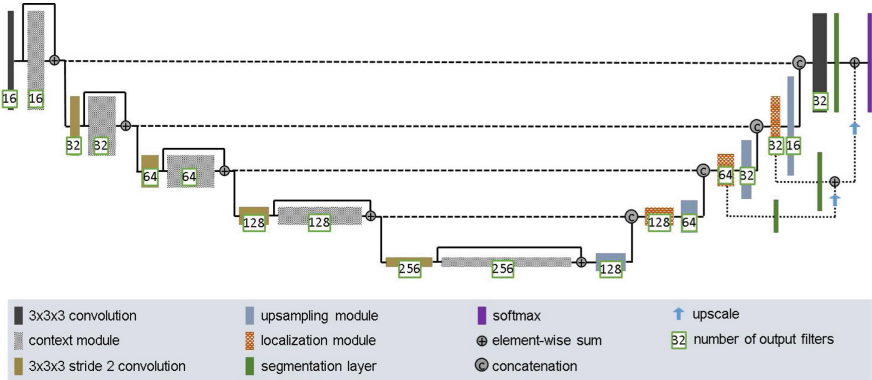


Fig. 3. Architecture of DFKZ Net

number of voxels for class k in patch. We set initial learning rate 5×10^{-4} and use instance normalization. We train the model for 90 epochs.

3D U-Net. U-Net [5, 14] is a classical network for biomedical image segmentation. It consists of a contracting path to capture context and a symmetric expanding path that enables precise localization with extension. Each pathway has three convolutional layers with dropout and pooling. And the contracting pathway and expanding pathway are linked by skip-connections. Each layer contains $3 \times 3 \times 3$ convolutional kernels. The first convolutional layer has 32 filters, while deeper layers contains twice filters than previous shallower layer.

For implementation, we use Adam optimizer [12], and we use instance normalization [15]. In addition, we utilize cross entropy as loss function. The initial learning rate is 0.001, the model is trained for 4 epochs.

Ensemble of Models. In order to enhance segmentation performance and reduce model variance. We use voting/majority rule to build an ensemble model. During training process, different models are trained separately. In the testing stage, each model independently predicts the class for each voxel, the final class is determined by majority rule.

2.4 Feature Extraction

Quantitative phenotypic features from MRI scans can reveal the characteristics of brain tumors. Based on the segmentation result, we extract radiomics features from edema, non-enhancing solid core and necrotic/cystic core and the whole tumor region respectively using *Pyradiomics* toolbox [16] (Fig. 4).

The modality used for feature extraction is depended on the intrinsic property of tumor subregion. For example, edema features are extracted from FLAIR

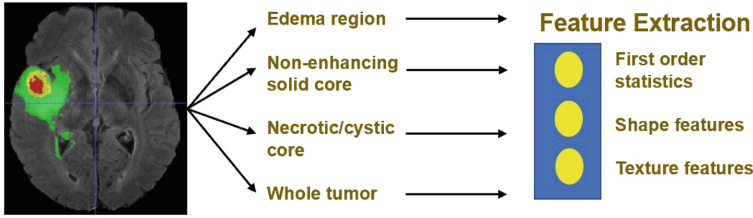


Fig. 4. Illustration of feature extraction

modality, since it is typically depicted by hyper-intense signal in FLAIR. Non-enhancing solid core features are extracted from T1c modality, since the appearance of the necrotic (NCR) and the non-enhancing (NET) tumor core is typically hypo-intense in T1-Gd when compared to T1. Necrotic/cystic core tumor features are extracted from T1c modality, since it is described by areas that show hyper-intensity in T1Gd when compared to T1.

The features we extracted can be grouped into three categories. The first category is first order statistics, which includes maximum intensity, minimum intensity, mean, median, 10th percentile, 90th percentile, standard deviation, variance of intensity value, energy, entropy and others. These features characterize the grey level intensity of tumor region.

The second category is shape features, which include volume, surface area, surface area to volume ratio, maximum 3D diameter, maximum 2D diameter for axial, coronal and sagittal plane respectively, major axis length, minor axis length and least axis length, sphericity, elongation and other features. These features characterize the shape of tumor region.

The third category is texture features, which include 22 grey level co-occurrence matrix (GLCM) features, 16 grey level run length matrix (GLRLM) features, 16 Grey level size zone matrix (GLSZM) features, five neighboring gray tone difference matrix (NGTDM) features and 14 grey level dependence matrix (GLDM) Features. These features characterize the texture of tumor region.

Not only do we extract features from original images, but we also extract features from Laplacian of Gaussian (LoG) filtered images and images generated by wavelet decomposition. Because LoG filtering can enhance the edge of images, possibly enhance the boundary of tumor, and wavelet decomposition can separate images into multiple levels of detail components (finer or coarser). More specifically, from each region, 1131 features are extracted, including 99 features extracted from the original image, and 344 features extracted from Laplacian of Gaussian filtered images, since we use 4 filters with sigma value 2.0, 3.0, 4.0, 5.0 respectively, and 688 features extracted from 8 wavelet decomposed images (all possible combinations of applying either a High or a Low pass filter in each of the three dimensions). In total, for each patient, we extract $1131 \times 4 = 4524$ radiomic features, these features are combined with clinical data (age and resection state) for survival prediction. The values of these features are normalized by subtracting the mean and scaling to unit variance.

2.5 Feature Selection

A portion of features we extracted are redundant or irrelevant to survival prediction. In order to enhance performance and reduce overfitting, we applied feature selection to select a subset of features that have the most predictive power. Feature selection is divided into two steps: importance ranking and cross validation. We rank the importance of features by fitting a decision tree regressor with gradient boosting using training data, then the importance of features can be determined by how effectively the feature can reduce intra-node standard deviation in leaf nodes. The second step is to select the optimal number of best features for prediction by cross validation. In the end, we select 14 features and their importance are listed as follows: (Abbreviations: wt = edema, tc = tumor core, et = enhancing tumor, full = whole tumor; The detailed feature definition can be found at <https://pyradiomics.readthedocs.io/en/latest/features.html>, last accessed on 30 June 2018) (Table 1).

Not surprisingly, age has the most predictive power among all features. The rest of features selected come from both original images and derived images. And we found that most features selected are come from images generated by wavelet decomposition.

2.6 Survival Prediction

Based on the 14 features selected, we trained a random forest regressor for final survival prediction. We set the number of base regressor as 100, and bootstrap samples when building trees.

Table 1. Selected most predicative features

| Extracted from | Name | Subregion | Score |
|--------------------|---|-----------|-------------|
| clinical | age | NA | 0.037375134 |
| wavelet-LHL | glcm_ClusterShade | wt | 0.036912293 |
| log-sigma-4.0mm-3D | glcm_Correlation | tc | 0.035558309 |
| log-sigma-2.0mm-3D | gldm_LargeDependenceHighGrayLevelEmphasis | tc | 0.026591038 |
| wavelet-LHL | glcm_Informational Measure of Correlation | et | 0.022911978 |
| wavelet-HLL | firstorder_Maximum | et | 0.020121927 |
| wavelet-LHL | firstorder_Skewness | et | 0.019402119 |
| original image | glcm_Autocorrelation | et | 0.014204463 |
| wavelet-HHH | gldm_LargeDependenceLowGrayLevelEmphasis | full | 0.014085406 |
| log-sigma-4.0mm-3D | firstorder_Mwtian | wt | 0.013031814 |
| wavelet-HLH | glcm_JointEntropy | wt | 0.013023534 |
| wavelet-LHH | glcm_ClusterShade | tc | 0.012335471 |
| wavelet-HLL | glszm_LargeAreaHighGrayLevelEmphasis | full | 0.011980896 |
| original image | firstorder_10Percentile | wt | 0.011803132 |

3 Experiments

3.1 Dataset

We utilize the BraTS 2018 dataset [1–4, 13] to evaluate the performance of our methods. The training set contains images from 285 patients, including 210 HGG and 75 LGG. The validation set contains MRI scans from 66 patients with brain tumors of unknown grade. The test set contains images from 191 patients with brain tumor, in which 77 patients have resection state of Gross Total Resection (GTR) and are evaluated for survival prediction. Each patient was scanned with four sequences: T1, T1c, T2 and FLAIR. All the images were skull-stripped and re-sampled to an isotropic 1 mm^3 resolution, and the four sequences of the same patient had been co-registered. The ground truth was obtained by manual segmentation results given by experts. Segmentation annotations comprise of the following tumor subtypes: Necrotic/non-enhancing tumor (NCR), peritumoral edema (ED), and Gd-enhancing tumor (ET). Resection status and patient age are also provided. The overall survival (OS) data, defined in days is also included in training set (Fig. 5).

3.2 Segmentation Result

We train the model using the 2018 MICCAI BraTS training set with methods described above. Then we applied the trained model for prediction on validation set and test set. We compared the segmentation result of ensemble model with individual model on validation set, the result demonstrates that the ensemble model performs better than individual models on enhancing tumor and whole tumor, while CA-CNN performs marginally better on tumor core (Table 2).

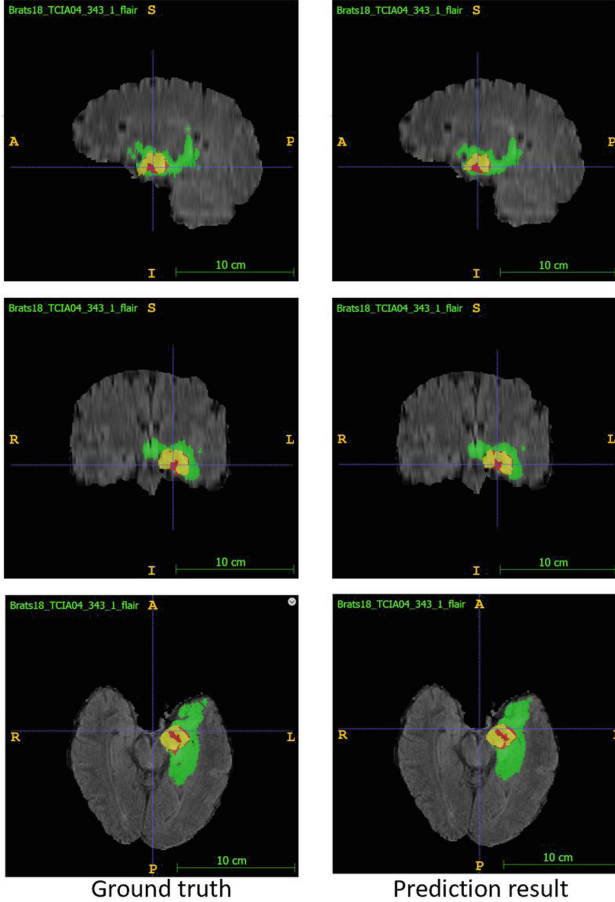
Table 2. Evaluation result of ensemble model and individual model

| Model | Enhancing tumor | Whole tumor | Tumor core |
|----------------|-----------------|----------------|----------------|
| CA-CNN | 0.77682 | 0.90282 | 0.85392 |
| DFKZ Net | 0.76759 | 0.89306 | 0.82459 |
| 3D U-Net | 0.78088 | 0.88762 | 0.82567 |
| Ensemble model | 0.80522 | 0.90944 | 0.84943 |

The predicted segmentation labels are uploaded to the CBICA’s Image Processing Portal (IPP) for evaluation. BraTS Challenge uses two schemes for evaluation: Dice score and the Hausdorff distance (95%). In test phase, we rank at 5th place out of 60+ teams. The evaluation result of segmentation on validation set and test set are listed as follows (Table 3).

Table 3. Evaluation result of ensemble model for segmentation

| Stage | Metric | Enhancing tumor | Whole tumor | Tumor core |
|------------|-----------------------|-----------------|-------------|------------|
| Validation | Mean Dice | 0.80522 | 0.90444 | 0.84943 |
| | Mean Hausdorff95 (mm) | 2.77719 | 6.32753 | 6.37318 |
| Test | Mean Dice | 0.71712 | 0.87615 | 0.79773 |
| | Mean Hausdorff95 (mm) | 4.97823 | 7.20086 | 6.47348 |

**Fig. 5.** Examples of segmentation result compared with ground truth Green: edema, Yellow: non-enhancing solid core, Red: enhancing core (Color figure online)

3.3 Survival Prediction Result

Based on the segmentation result of brain tumor subregions, we extract features from brain tumor sub-regions segmented from MRI scans and trained the survival prediction model as described above. Then we use the model to predict

patient’s overall survival on validation set and test set. The predicted overall survival is uploaded to the IPP for evaluation. We use two schemes for evaluation: classification of subjects as long-survivors (>15 months), short-survivors (<10 months), and mid-survivors (between 10 and 15 months) and median error (in days). In test phase, we rank at second place out of 60+ teams. The evaluation result is listed as follows (Table 4).

Table 4. Evaluation result of survival prediction

| Stage | Classification accuracy | Median error |
|------------|-------------------------|--------------|
| Validation | 46.4% | 217.92 |
| Test | 61.0% | 181.37 |

4 Conclusion

In this paper, we present an automatic framework for prediction of survival in glioma using multimodal MRI scans and clinical features. Firstly deep convolutional neural network (CNN) is used to segment tumor region from MRI scans, then radiomics features are extracted and combined with clinical features to predict overall survival. For tumor segmentation, we use ensembles of three different 3D CNN architectures for robust performance through voting/majority rule. This approach can effectively reduce model bias and boost performance. For survival prediction, we extract shape features, first order statistics and texture features from segmented tumor sub-region, then use decision tree and cross validation to select features. Finally, a random forest model is trained to predict the overall survival of patients. On 2018 MICCAI BraTS Challenge, our method ranks at second place and 5th place out of 60+ participating teams on survival prediction task and segmentation task respectively, achieving a promising 61.0% accuracy on classification of long-survivors, mid-survivors and short-survivors. In the future, we will explore different network architectures and training strategies to further improve our result. We will also design new features and optimize our feature selection methods for survival prediction.

References

1. Bakas, S., et al.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection. *Cancer Imaging Arch.* **286** (2017)
2. Bakas, S., et al.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection. *Cancer Imaging Arch.* **286** (2017)
3. Bakas, S., et al.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci. Data* **4**, 170117 (2017)
4. Bakas, S., Reyes, M., 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](https://arxiv.org/abs/1811.02629)* (2018)

5. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 424–432. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_49
6. Ferlay, J., Shin, H.R., Bray, F., Forman, D., Mathers, C., Parkin, D.M.: Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer* **127**(12), 2893–2917 (2010)
7. Gillies, R.J., Kinahan, P.E., Hricak, H.: Radiomics: images are more than pictures, they are data. *Radiology* **278**(2), 563–577 (2016)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
9. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv preprint [arXiv:1502.03167](https://arxiv.org/abs/1502.03167) (2015)
10. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H.: Brain tumor segmentation and radiomics survival prediction: contribution to the BRATS 2017 challenge. In: Crimi, A., Bakas, S., Kuijff, H., Menze, B., Reyes, M. (eds.) BrainLes 2017. LNCS, vol. 10670, pp. 287–297. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75238-9_25
11. Kamnitsas, K., et al.: Ensembles of multiple models and architectures for robust brain tumour segmentation. In: Crimi, A., Bakas, S., Kuijff, H., Menze, B., Reyes, M. (eds.) BrainLes 2017. LNCS, vol. 10670, pp. 450–462. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75238-9_38
12. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. In: International Conference on Learning Representations (2015)
13. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* **34**(10), 1993 (2015)
14. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
15. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Instance normalization: the missing ingredient for fast stylization. arXiv 2016. arXiv preprint [arXiv:1607.08022](https://arxiv.org/abs/1607.08022)
16. Van Griethuysen, J.J.M., et al.: Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* **77**(21), e104–e107 (2017)
17. Wang, G., Li, W., Ourselin, S., Vercauteren, T.: Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In: Crimi, A., Bakas, S., Kuijff, H., Menze, B., Reyes, M. (eds.) BrainLes 2017. LNCS, vol. 10670, pp. 178–190. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75238-9_16