Computer Vision Radiology

William Bakst, Linus Meyer-Teruel, Cameron Andrews

Motivation

- Cancer was responsible for 600,000 deaths in the United States in 2016, and is one of the leading causes of death in the world today.
- Early detection is the most effective tool we have to stymie this horrible disease, and with the rise and refinement of deep learning methods coupled with the dramatic increase in the power and accessibility of computational resources, machines seem perfectly positioned to alleviate some of the glaring inefficiencies and shortcomings of the diagnostic status quo.

Datasets

- We are using the BraTS 2015 Dataset, which includes multiple imaging modalities (T1/T2 MRI tissue contrasts, T2 FLAIR, and T1 contrast-enhanced MRI), and variety of presenting phenotypes (primary/secondary tumors, solid and infiltrative growing tumor profiles)
- In total, the data consists of 274 training MR images, of which 220 are high grade gliomas and 54 are low grade gliomas. Along with ground-truth pixel labels (on/off) for each scan.
- The dataset is broken up into two segments: the data itself (skull-stripped brain MRIs) and the ground-truth segmentation of the data, hand-annotated by radiologists.

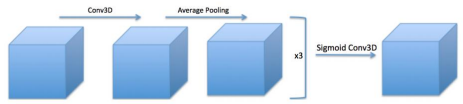
Challenges

- The manipulation and interpretation of our image data, primarily due to fuzzy imagery and small datasets.
- MRI data is recorded in three dimensions, and requires a network architecture that can properly process it as such. In addition, 3D imaging modalities are more prone to blurring and image quality degeneration as the result of patients moving around on the stage during the course of their 30-40 minute scan.
- Accounting for this in a scalable fashion can be difficult. We used data augmentation and tried different architectures to help combat this.

Models

Baseline Model

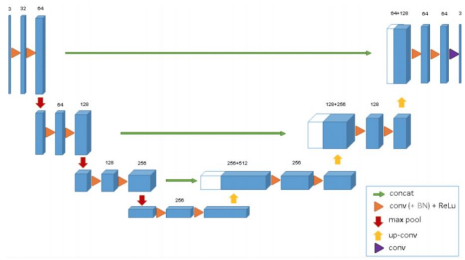
- We used a four layer 3D convolutional neural network



UNet-3D

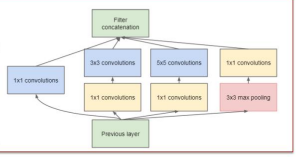
- Our UNet-3D model is built based on the original Unet-3D paper. We use the same number of standard convolutional layers, maxpools, and upconvolutions for a total of 23 layers and 19 million parameters. Before each max-pooling, we have a double convolutional layer of the form:

  CONV->ReLU->BATCHNORM->CONV->ReLU->BATCHNORM.



UNet-3D with Inception Layers

- For this model we have altered the original UNet-3D architecture by replacing the double convolutional layers with single inception layers. This increases the total convolutions to 42 while keeping the number of maxpools and upconvolutions constant but uses only approximately 2 million parameters.



References

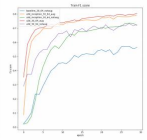
Menze et al, The Multimodal Brain TumorImage Segmentation Benchmark (BRATS), IEEE Trans. Med. Imaging, 2015

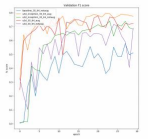
Cicek et al; 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation 2016. arXiv:1606.06650

Szegedy et al; Going Deeper with Convolutions, 2014. arXiv:1409.4842

Experimental Results

Model	Train Loss	Dev Loss	Train F1	Dev F1	Params
Baseline	0.037	0.0033	0.572	0.579	69,729
U3D	0.0021	0.0022	0.734	0.716	2,614,949
U3D-Aug	0.0013	0.0016	0.805	0.778	2,614,949
U3D-Inception	0.0019	0.0019	0.736	0.741	19,078,337
U3D-Inception Aug	0.0014	0.0017	0.793	0.797	19,078,337





Discussion

- Our networks performed very well on the BRATS 2015 dataset. We attribute this partly to the relative simplicity of the diagnostic task (as glioblastomas presents quite prominently relative to most other disease states).
- Our data was also limited in scale, and the ground-truth values for the training set were derived computationally and verified manually. While we are confident in the efficacy of our models, we want to extend our training set to include more data that was manually annotated by doctors in order to improve the generalizability of our models.

Future Work

- A handful of 3D, fully convolutional architectures have done well on the diagnosis and segmentation of everything from brain tumors to multiple sclerosis. There are many more models that we would still like to build, evaluate, and compare to our current models. For future work, we would do some follow-up research on a handful of architectures such as SegNet, FCN, Nabla-Net, DeepMedic, and others.
- We also plan to evaluate our models on different, similar datasets such as the new BRATS 2017 dataset, which was hand-labeled by a committee of doctors, as well as a proprietary dataset that we have recently been given access to. We believe that our models will be able to perform quite well on these datasets given the performance we have seen on our current dataset.