

Fully Convolutional U-Net Approaches to Brain Tumor Image Segmentation

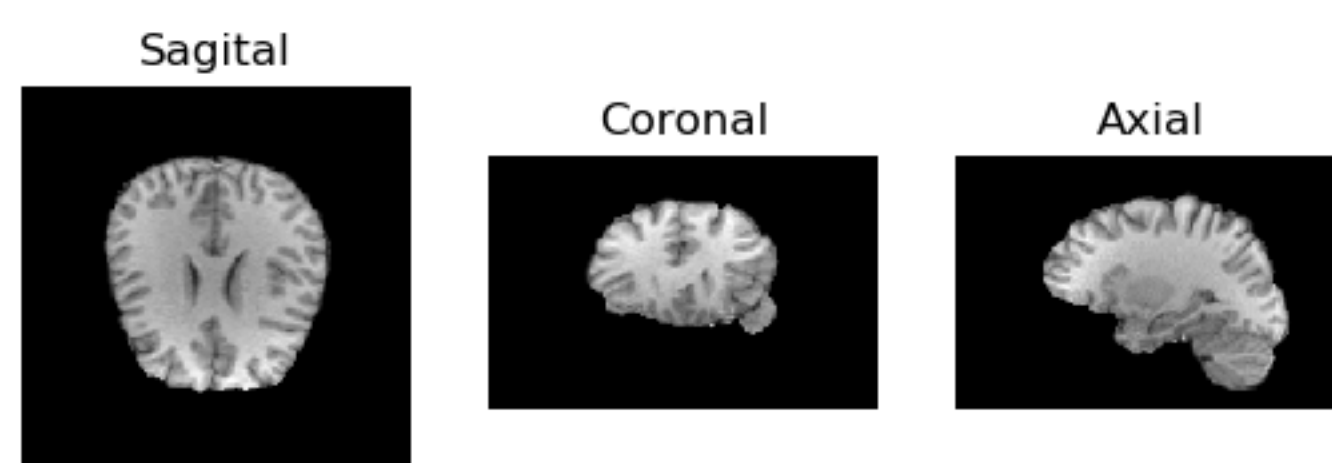
Cameron Backes, Jonathan Deaton
cbackes@stanford.edu, jdeaton@stanford.edu

Introduction

Noninvasive methods of brain imaging, most commonly Magnetic Resonance Imaging (MRI), are routinely used to identify and locate tumors in the brain; however, this procedure is time consuming and must be performed manually by medical professionals. As such, there is significant societal value in applying newly emerging deep learning approaches to automate and improve the accuracy of segmentation for a variety of medical applications, particularly for deployment in regions where medical personnel are sparse. We have created, trained, and tested three models using a 3D Fully Convolutional U-Net architecture to perform segmentation of brain tumors in MR images, as a part of the 2018 Brain Tumor Segmentation (BraTS) challenge. Our best model achieved an average Dice coefficient of 0.87 on our test set, making our model's performance on par with state-of-the-art performances from recent years' BraTS challenges.

Dataset and Features

The 2018 Brain Tumor Segmentation (BraTS) dataset was collected by medical professionals from numerous institutions, and consists of 3D MRI brain scans from 285 individuals with brain tumors, along with age and survival information for each individual, and ground-truth tumor segmentation labels for tumor voxels manually-revised by expert board-certified neuroradiologists.



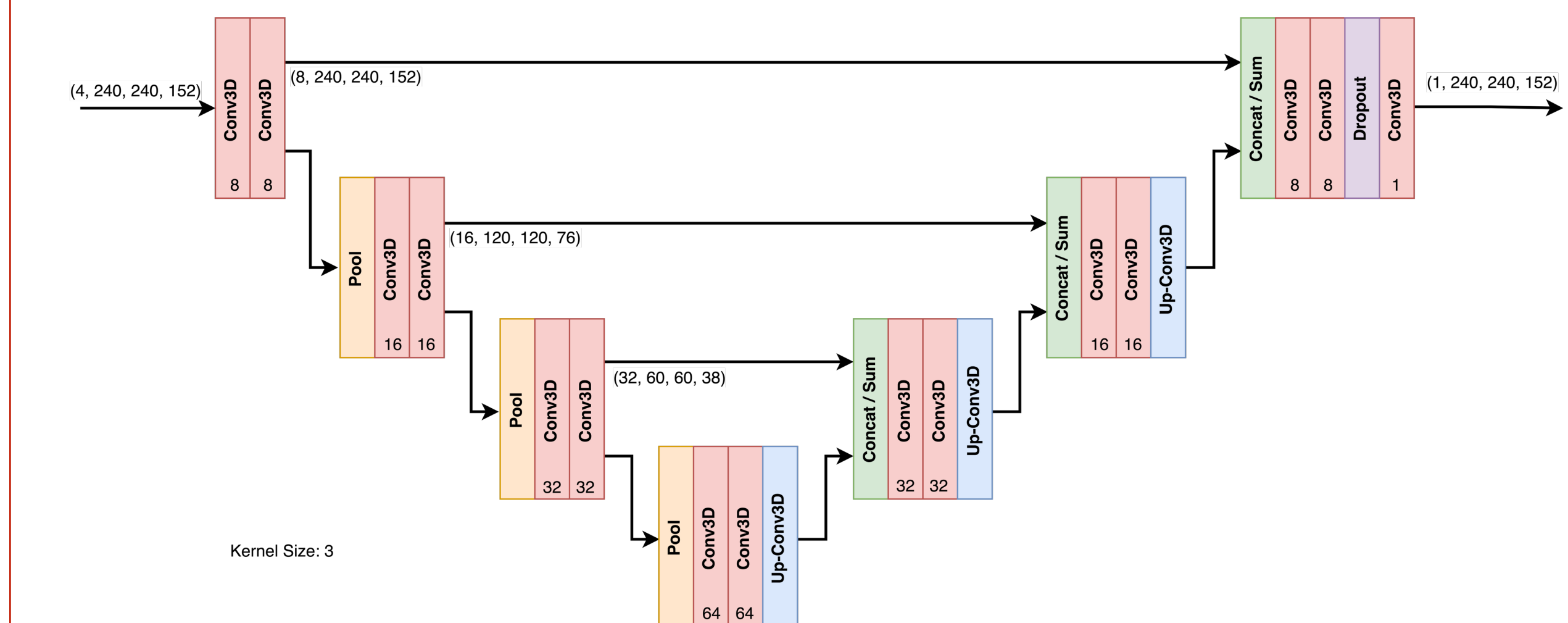
Each MRI scan corresponds to a single individual and consists of 4 different image types: T1, T1-enhanced, T2, and T2-FLAIR, and for each scan type there is a 240x240x155 voxel 3D gray-scale image input, where each voxel represents a feature in our model. We have augmented our training dataset by flipping each image along all 3 axes.

Models

We trained, and tested 3 models each of which utilizes a 3D extrapolation of the U-Net architecture, and which contain concatenations, element wise summations, and neither concatenation nor summation of corresponding down-sampling and up-sampling blocks, respectively. Each model uses a Dice coefficient loss function that is calculated given model segmentation X and ground-truth segmentation Y, as:

$$Dice = \frac{2 * |X \cap Y|}{|X| + |Y|}$$

Model Architecture:

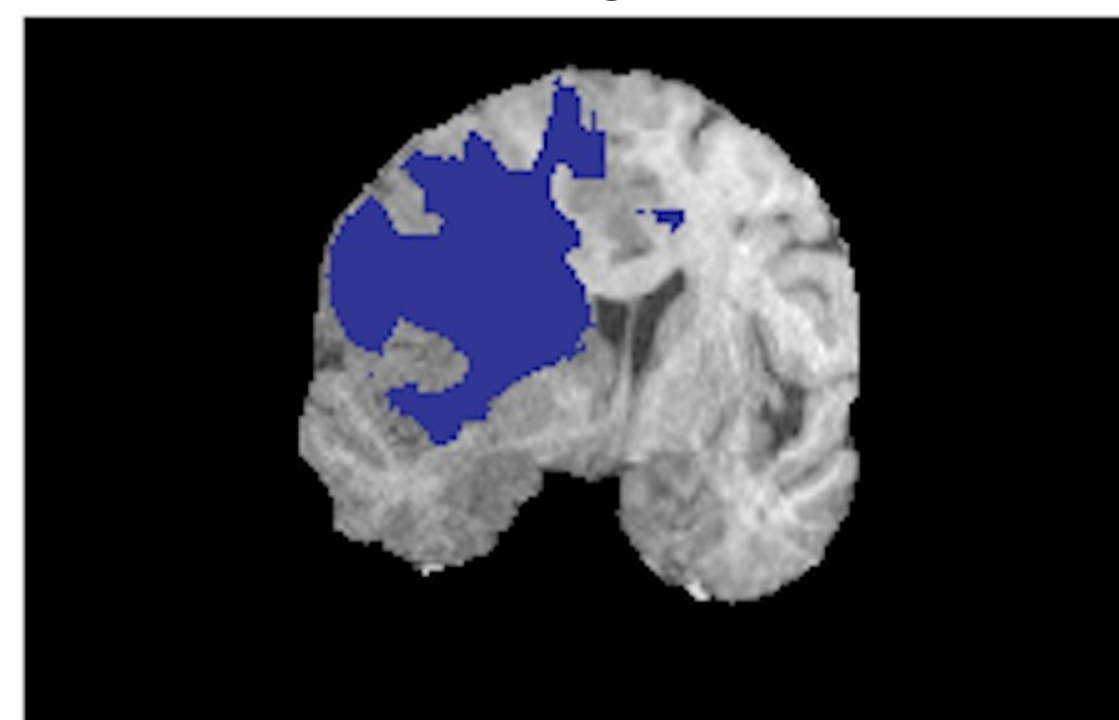


UNet Model	Weight initialization	Dropout	Batch Norm	Skip Connection	Conv Activation	Output Activation	Optimization
Concatenation	Xavier-He	FALSE	TRUE	Concatenation	Relu	Sigmoid	Adam
Dropout & Summation	Xavier-He	TRUE	TRUE	Summation	Relu	Sigmoid	Adam
Dropout & No Skip	Xavier-He	TRUE	TRUE	None	Relu	Sigmoid	Adam

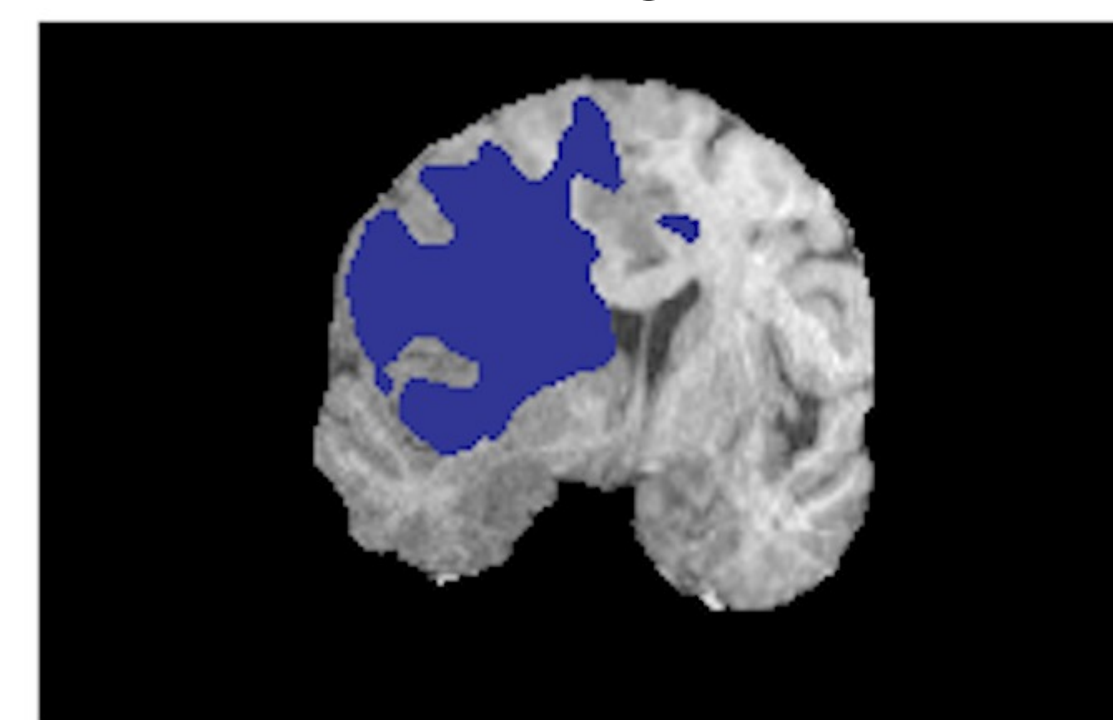
Results

Our best performing model was a 3D Fully Convolutional U-Net architecture that used concatenations as skip connections, which achieved 0.87 average dice score on the test set. However, our model which incorporated dropout and summation as opposed to concatenation achieved a nearly equivalent test dice score of .85, with a substantially reduced computational load. Finally, our model with neither concatenation nor summation performed substantially worse on both validation and test sets.

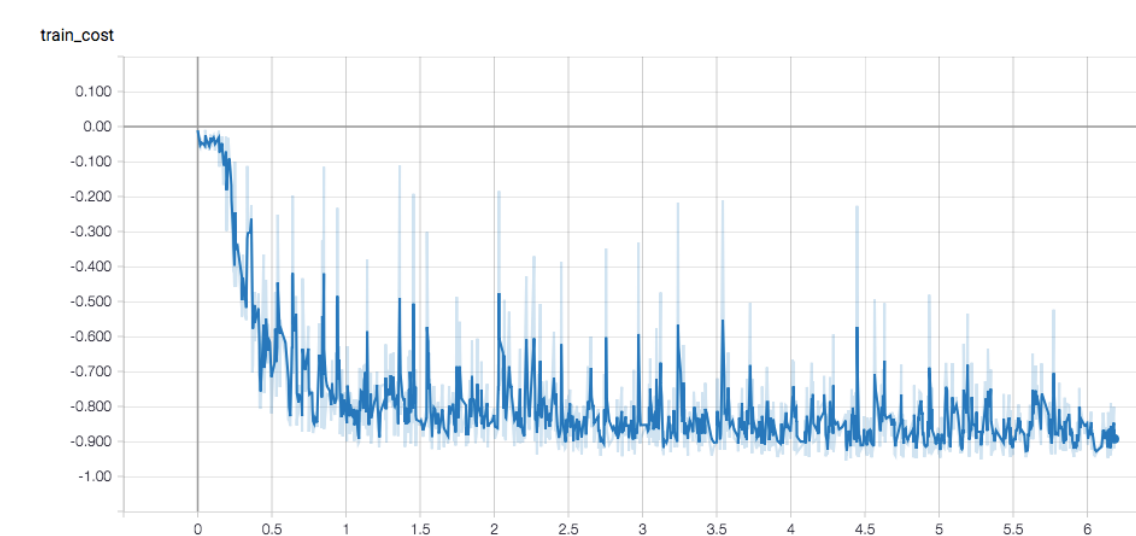
Model Segmentation



Ground-Truth Segmentation



Train Cost vs. Epoch



Dice Scores by Trial Class

UNet Model	Train (N = 204)	Validation (N = 40)	Test (N = 40)
Concatenation	0.91 +/- 0.047	0.89 +/- 0.07	0.87 +/- 0.072
Dropout & Summation	0.89 +/- 0.05	0.86 +/- 0.08	0.85 +/- 0.09
Dropout & No Skip	0.82 +/- 0.14	0.77 +/- 0.25	0.81 +/- 0.12

Discussion

We explored the application of a 3D Fully Convolutional U-Net architecture with and without concatenations for brain tumor image segmentation, and demonstrated high, generalizable performance is achievable in both cases. We are pleasantly surprised to report that our best model, which uses concatenation, achieves near state-of-the-art performance. We believe this model performs best because skip connections allow for spatial information to be propagated most effectively across layers. However, building off of existing literature, we showed that summations can be used as an alternative to concatenations and result in nearly equivalent performance. Summations also substantially reduced computational cost of training given limited GPU memory. We also found that removing skip connections altogether decreased model performance drastically despite the reduced memory footprint during training.

Future Directions

Training our model with randomly chosen tumor-containing patches from MRIs rather than full MRIs would significantly reduce the computational cost of training. To improve our model's dice score, desirable next steps are augmenting the dataset with MRI scans from previous years' BraTS challenges, and incorporating a post processing step in which our model's output parameters are passed into and tuned by a conditional random field (CRF).

References

- B. Menze, M. Reyes, K. Van Leemput, "The multimodal brain tumor image segmentation benchmark (BRATS)", IEEE Trans. Med. Imag., vol. 34, no. 10, pp. 1993-2024, Oct. 2015.
- O. Cicek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation, 2016.