

# A Fully Automated Deep Learning Network for Brain Tumor Segmentation

Chandan Ganesh Bangalore Yogananda<sup>1</sup>, Bhavya R. Shah<sup>1</sup>, Maryam Vejdani-Jahromi<sup>1</sup>, Sahil S. Nalawade<sup>1</sup>, Gowtham K. Murugesan<sup>1</sup>, Frank F. Yu<sup>1</sup>, Marco C. Pinho<sup>1</sup>, Benjamin C. Wagner<sup>1</sup>, Kyrre E. Emblem<sup>2</sup>, Atle Bjørnerud<sup>3</sup>, Baowei Fei<sup>4</sup>, Ananth J. Madhuranthakam<sup>1</sup>, and Joseph A. Maldjian<sup>1</sup>

<sup>1</sup>Department of Radiology, Advanced Neuroscience Imaging Research Lab (ANSIR Lab), University of Texas Southwestern Medical Center, Dallas, TX; <sup>2</sup>Department of Diagnostic Physics, Division of Radiology and Nuclear Medicine, Oslo University Hospital, Oslo, Norway; <sup>3</sup>Computational Radiology and Artificial Intelligence (CRAI), Division of Radiology and Nuclear Medicine, Oslo University Hospital, Oslo, Norway; and <sup>4</sup>Department of Bioengineering, The University of Texas at Dallas, Richardson, TX

## Corresponding Author:

Joseph A. Maldjian, MD  
Department of Radiology, UT Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, Texas 75390-9178;  
E-mail: Joseph.Maldjian@utsouthwestern.edu

**Key Words:** Brain tumor segmentation, deep learning, BraTS, machine learning, CNN (convolutional neural networks), MRI, Dense UNet

**Abbreviations:** brain tumor segmentation (BraTS), whole tumor (WT), tumor core (TC), enhancing tumor (ET), nonenhancing tumor including necrosis (NEN), edema (ED), magnetic resonance imaging (MRI), convolutional neural networks (CNNs), high-grade glioma (HGG), low-grade glioma (LGG), multi volume fusion (MVF)

## ABSTRACT

We developed a fully automated method for brain tumor segmentation using deep learning; 285 brain tumor cases with multiparametric magnetic resonance images from the BraTS2018 data set were used. We designed 3 separate 3D-Dense-UNets to simplify the complex multiclass segmentation problem into individual binary-segmentation problems for each subcomponent. We implemented a 3-fold cross-validation to generalize the network's performance. The mean cross-validation Dice-scores for whole tumor (WT), tumor core (TC), and enhancing tumor (ET) segmentations were 0.92, 0.84, and 0.80, respectively. We then retrained the individual binary-segmentation networks using 265 of the 285 cases, with 20 cases held-out for testing. We also tested the network on 46 cases from the BraTS2017 validation data set, 66 cases from the BraTS2018 validation data set, and 52 cases from an independent clinical data set. The average Dice-scores for WT, TC, and ET were 0.90, 0.84, and 0.80, respectively, on the 20 held-out testing cases. The average Dice-scores for WT, TC, and ET on the BraTS2017 validation data set, the BraTS2018 validation data set, and the clinical data set were as follows: 0.90, 0.80, and 0.78; 0.90, 0.82, and 0.80; and 0.85, 0.80, and 0.77, respectively. A fully automated deep learning method was developed to segment brain tumors into their subcomponents, which achieved high prediction accuracy on the BraTS data set and on the independent clinical data set. This method is promising for implementation into a clinical workflow.

## INTRODUCTION

Gliomas are the most common primary brain malignancy and represent a heterogeneous set of tumors. Gliomas are classically divided between high- and low-grade tumors based on their histopathology, immunohistochemistry, and genetic profiles. Glioma segmentation based on magnetic resonance imaging (MRI) can be useful in predicting aggressiveness and response to therapy. Currently, MRI segmentation of gliomas is largely based on imaging correlates of histopathological findings. Gliomas are generally segmented into active tumor, necrotic tissue, and surrounding edema (ED) based on conventional MRI sequences. Accurate image-based segmentation depends on the ability to differentiate the MRI signal of these subcomponents. Multiple MRI sequences that generate variable tissue contrast are simultaneously used for glioma segmentation (1). Manual tumor segmentation is a tedious,

time-intensive task that requires a human expert to delineate components. Therefore, manual tumor segmentation is often fraught with intra-rater and inter-rater variability, resulting in imprecise boundary demarcation (2, 3). An intra-rater variability of 20% and an inter-rater variability of 28% has been reported for manual segmentation of brain tumors (4, 5).

To address these shortcomings, automated machine learning algorithms have been developed to segment gliomas. Machine learning algorithms have been shown to improve glioma segmentation by decreasing variability and the time required for manual segmentation (2, 3, 6). Glioma segmentation is essentially a voxel-level classification task. Algorithms for voxel-level classification can be broadly divided into classic machine learning techniques, such as support vector machines, and deep learning methods, such as convolutional neural networks (CNNs).

CNN-based methods have been shown to outperform classic machine learning methods (3, 7–12).

The Multimodal Brain Tumor Image Segmentation Benchmark (BraTS) challenge was established in 2012 to gauge and facilitate the progress of automated glioma segmentation (4, 13). The BraTS data set represents a valuable, publicly available data set for developing and evaluating tumor segmentation algorithms. The BraTS data set consists of multiparametric magnetic resonance (MR) scans of patients with low- and high-grade glioma that have been manually segmented by expert raters. The BraTS data are provided with the following 3 ground truth labels including (1) enhancing tumor (ET), (2) nonenhancing tumor including necrosis (NEN), and (3) ED. Various algorithms in the BraTS challenge are evaluated based on label outputs of whole tumor (WT), tumor core (TC), and ET. WT consists of enhancing components, nonenhancing components including necrosis, and edema (ET + NEN + ED). TC consists of enhancing components and nonenhancing components including necrosis (ET + NEN). Enhancing tumor consists of just the enhancing component. To evaluate an algorithm's performance, data can be uploaded to the BraTS validation server that reports DICE coefficients for TC, WT, and ET (14).

For the 2017 BraTS challenge, there were 46 cases provided on the BraTS validation data set. In 2018, the validation set was expanded to 66 cases, including the previous 46. The training data set remained the same between 2017 and 2018. As such, the BraTS 2018 data set allows developers to compare results with the top performers from 2017 and 2018. The BraTS validation server calculates DICE scores for WT, TC, and ET, but it does not provide DICE scores for ED and NEN. Although ED and NEN DICE scores are not reported by the BraTS server, they can be of value in comparing algorithmic performance. The BraTS challenge also evaluates performance on a held-out test data set; however, these data are available for only a brief time period during the challenge.

We developed a 3D Dense UNet CNN for glioma segmentation that can be easily incorporated into clinical workflow. The algorithm's performance was evaluated using the BraTS validation server for WT, TC, and ET. The algorithm was also tested on an independent clinical data set from Oslo University Hospital, Oslo, Norway, and the DICE scores for WT, TC, and ET are reported using expert segmentation as the ground truth. In addition, we also report DICE scores for ED and NEN for both data sets.

## MATERIALS AND METHODS

### BraTS Data Set

Multiparametric MRI data (T2w, T2w-FLAIR, T1, and T1 postcontrast) were obtained from the BraTS2018 data set. The BraTS2018 data set consisted of a total of 285 subjects: 210 subjects with high-grade glioma (HGG) and 75 subjects with low-grade glioma (LGG) (4, 5). The data set included 3 ground truth labels for (1) ET, (2) NEN, and (3) ED. The BraTS data were already reoriented to LPS (left posterior-superior) coordinate system, coregistered to T1c, registered to the SR124 template and resampled to 1-mm, and skull stripped. The images were N4bias corrected (13, 15) to remove the RF inhomogeneities and intensities were normalized to zero mean and unit variance before using the data.

### Oslo Data Set

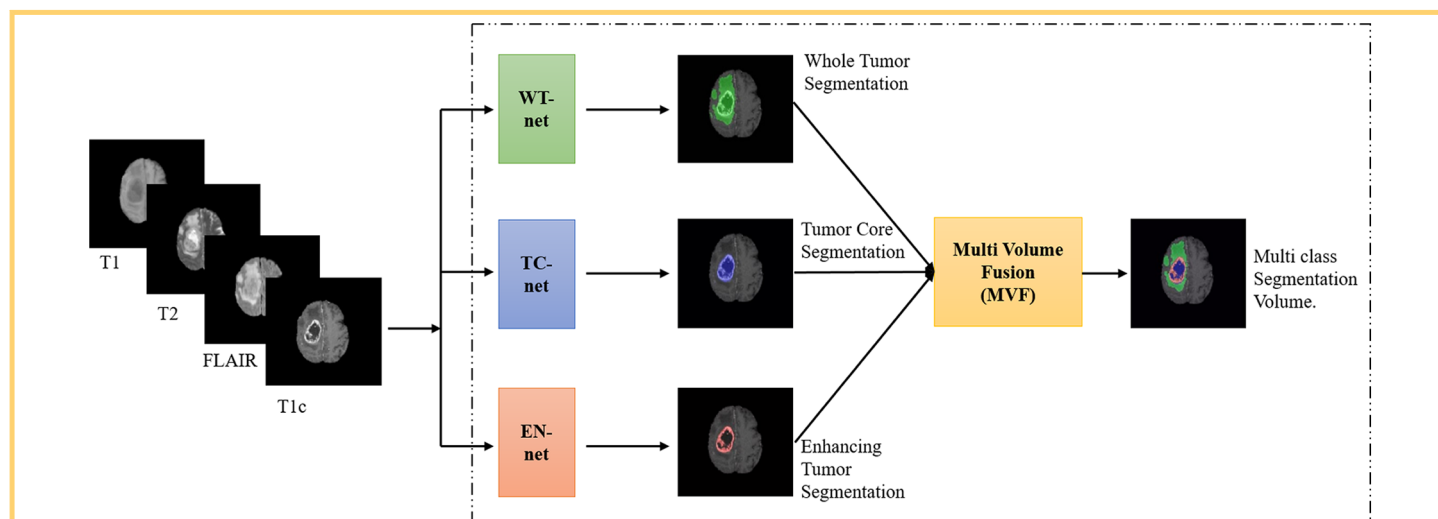
Multiparametric MRI data (T2w, T2w-FLAIR, T1, and T1c/T1 postcontrast) were obtained from Oslo University Hospital (16, 17). The Oslo data set consisted of 52 preoperative subjects with LGG and HGG (age > 18 years) scanned from 2003 to 2012. Most images were acquired with anisotropic voxels, typically used in routine clinical 2D images, characterized by high in-plane resolution (0.7–0.9 mm) and low through-plane resolution (slice thickness varying between 5 and 6 mm).

The original DICOM images were converted to NifTI format for ease of processing. The Oslo data set was manually segmented by an in-house neuroradiologist for the same 3 labels used in the BraTS data set (ED, NEN, and ET). Images from the Oslo data set were preprocessed following the same steps used in the BraTS data set. The preprocessing pipeline was developed using ANTS (18), and included coregistration to the T1 postcontrast, registering to the SR124 template (13, 15), resampling to 1 mm<sup>3</sup> isotropic resolution, skull stripping, N4BiasCorrection (19), and intensity normalization to zero-mean and unit variance.

### Network Architecture

The histologic complexity of gliomas poses a challenge to automated tumor segmentation methods. To simplify the segmentation problem, a triple network architecture was designed (Figure 1). Each model was trained separately to predict WT (WT-net), TC (TC-net), and ET (EN-net) as a binary task. The networks used a 3D patch-based approach. Multiparametric images were passed through the Dense UNet (Figure 2A). The initial convolution generated 64 feature maps that were subsequently used to create dense blocks. Each dense block consisted of 5 layers (Figure 2B). Each layer included the following 4 sequentially connected sub-layers: (1) batch normalization, (2) rectified linear unit (ReLU), (3) 3D convolution, and (4) 3D spatial dropout. The first layer in dense block 1 had 32 feature maps as its input. At each layer, the input was used to generate *k* feature maps, which were then concatenated to the next layer input; this was then applied to create another *k* feature map. To generate the final dense block output, inputs from each layer were concatenated with the output of the last layer. At the end of each dense block, the input to the dense block was also concatenated to the output of that dense block. The output of each dense block followed a skip connection to the adjacent decoder part. In addition, each dense block output went through a transition down block until the bottle neck block (Figure 2C). With this connecting pattern, all feature maps were reused such that every layer in the architecture received a direct supervision signal (20). On the decoder side, a transition-up block preceded each dense block until the final convolution layer, which was followed by a sigmoid activation layer.

To preserve a high number of convolution layers and fit the complex model into GPU memory, 3 additional steps were used. (1) If a layer generated feature maps exceeding the initial number of convolution feature maps, then it was reduced to one-fourth of the total number of feature maps generated by that layer. (2) The total number of feature maps generated at the end of every dense block was reduced by a compression factor of 0.75. (3) A bottle neck block (dense block 4 in Figure 2A) was used to connect the encoder part of the network to the decoder part of the network. This bottle neck block reduced the feature maps



**Figure 1.** Schematic representation of the developed algorithm. The input images included T1, T2, T2-FLAIR, and T1 post-contrast (T1c). Whole tumor (WT-net) segments the WT, tumor core (TC-net) segments the TC, and enhancing tumor (EN-net) segments the ET. The output segmented volumes from each of these networks are combined using a triple volume fusion to generate a multiclass segmentation volume.

generated by the encoder part of the network by the same compression factor of 0.75. Owing to the large number of high-resolution feature-maps, a patch-based 3D Dense UNet approach was implemented. However, the higher resolution information was passed through the standard skip connections.

### Cross-Validation

To generalize the network’s performance and evaluate its reliability, a 3-fold cross validation was performed on the BraTS2018 data set (210 HGG and 75 LGG subjects). The data were randomly shuffled and equally split into 3 groups as training, in-training validation, and held-out testing (70 HGG and 25 LGG cases in each group). The in-training validation data set is used by the algorithm to test performance after each round of training, and update model parameters. Each fold of the cross-validation procedure represents a new training phase on a unique combination of the 3 groups. Network performance is only reported on the held-out testing group for each fold.

### Training

The 3-fold cross-validation procedure uses a relatively small sample of cases from the BraTS2018 data set for training each fold (95 cases). Before evaluation of the independent data set, the networks were retrained on a larger sample of the BraTS2018 data set using ~70% of the cases for training (200 cases including 150 HGG and 50 LGG), ~20% for in-training validation (65 cases including 48 HGG and 17 LGG) and ~10% (20 cases including 12 HGG and 8 LGG) held out for testing. Further, 75% overlapping patches were extracted from multiparametric brain MR images that had at least 1 nonzero pixel on the corresponding ground truth patch. Subsequently, 20% of patches were used for in-training validation. Data augmentation steps included horizontal flipping, vertical flipping, random rotation, and translational rotation. Downsampled data (128 × 128 × 128) was also

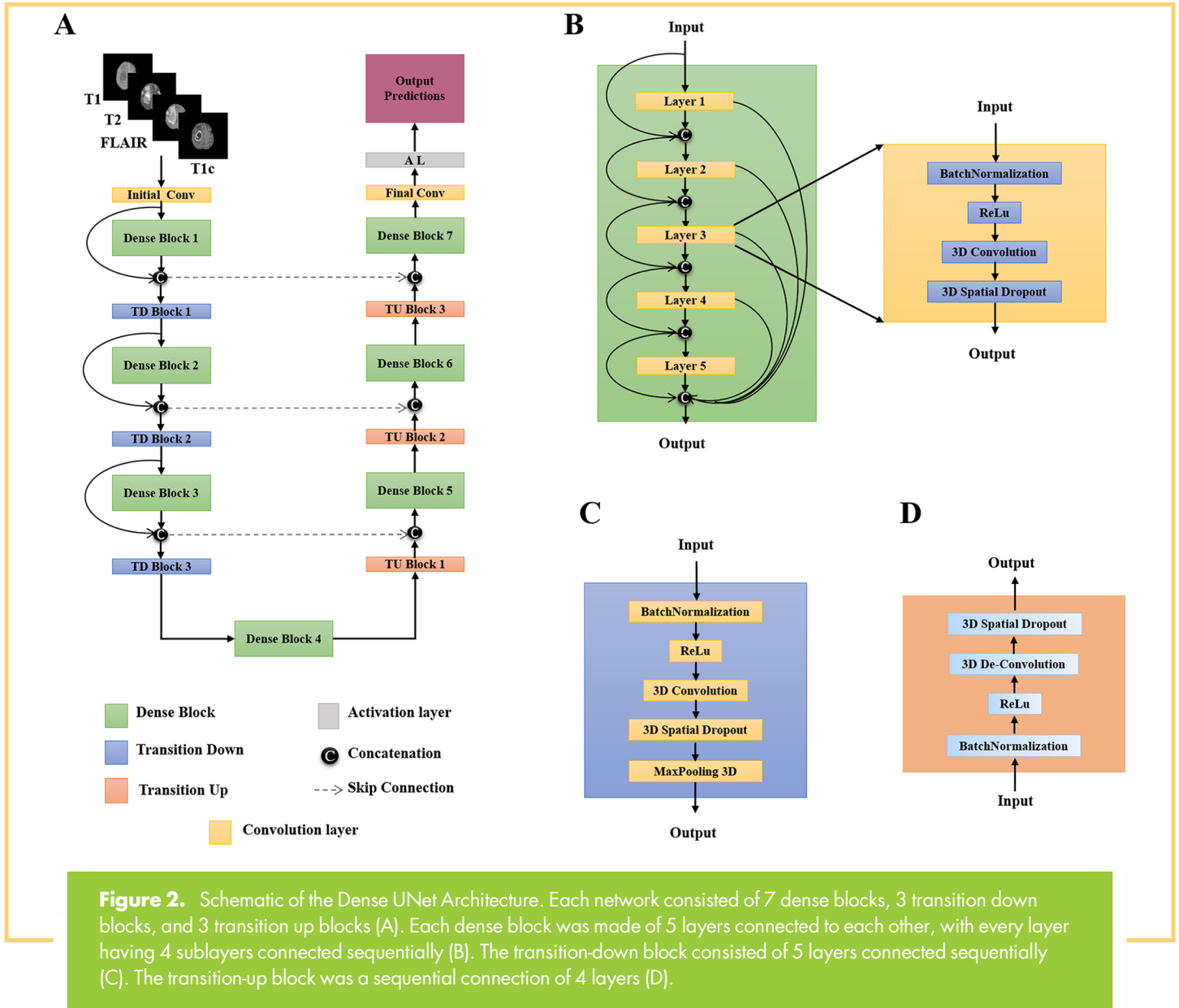
provided in the training as an additional data augmentation step. To eliminate the data leakage problem, no patch from the same subject was mixed in training, validation, or testing (21, 22). Labels of ED, ET, and NEN were fused to create a WT mask to train WT-net. A TC mask was created by fusing the labels of ET and NEN to train TC-net. ET labels were used separately to train EN-net. The networks were trained using Tensorflow (23), the Keras (24) python package, and Pycharm IDEs with adaptive moment estimation (Adam) (25) as the optimizer. The initial learning rate was set to  $10^{-5}$  with a batch size of 4 and maximal iteration of 100. Training was implemented on Tesla P100, P40 or K80 NVIDIA-GPUs.

### Testing

The final network was tested on the 20 held-out cases. Patches of  $32 \times 32 \times 32$  were extracted and provided to the network for testing. All of the prediction patches were then reconstructed to obtain a full segmentation volume. After obtaining the 3 separate segmentation output volumes from the 3 networks, they were fused in 2 steps. First, a 3D-connected components algorithm was applied to the WT-net output to generate a WT mask. Next the outputs from TC-net and EN-net were multiplied by the output from WT-net. This procedure, referred to as multivolume fusion (MVF), was designed to improve the prediction accuracy by removing false positives. The final network was also tested on 46 cases from BraTS2017 validation data set, 66 cases from BRATS 2018 validation data set, and 52 cases from the Oslo data set without any fine-tuning.

### Statistical Methods

The performance of each network was evaluated using the Dice coefficient (15), which determines the amount of spatial overlap between the ground truth segmentation (X) and the network segmentation (Y), as follows:



$$Dice = \frac{2|X_1 \cap Y_1|}{|X_1| + |Y_1|}$$

These output labels are defined as follows:

WT = edema + enhancing Tumor + nonenhancing tumor + necrosis

TC = enhancing tumor + nonenhancing tumor + necrosis

ET = enhancing tumor

Dice coefficients were also computed for:

Edema = WT – TC

Nonenhancing tumor and necrosis = TC – ET

## RESULTS

### Cross-Validation

Average Dice-scores for the 3-fold cross validation using 75% overlapping patches were 0.90, 0.82, and 0.79 for WT, TC, and

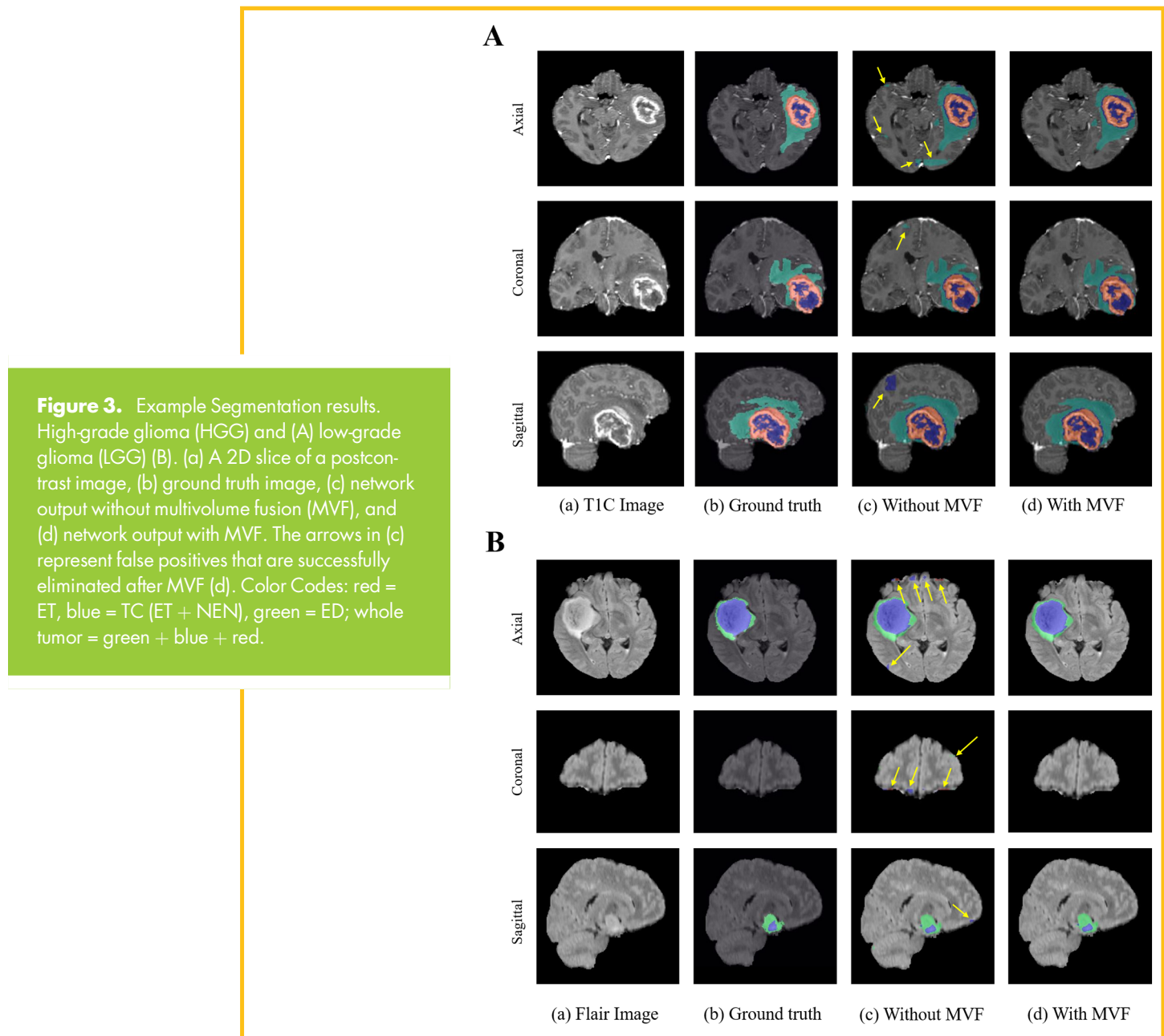
ET, respectively. Average Dice-scores for the 3-fold cross-validation using 85% overlapping patches were 0.92, 0.84, and 0.80 for WT, TC, and ET, respectively. Detailed Dice-scores are provided in the online supplemental data.

### Testing on 20 Held-Out Cases from BraTS2018

The network achieved Dice scores of 0.90, 0.84 and 0.80 with Hausdorff distance of 3.9 mm, 5.9 mm and 3.5 mm for WT, TC, and ET, respectively, on the 20 held-out cases (Figure 3). Sensitivities were 0.91, 0.85, and 0.81 for WT, TC, and ET, respectively, with 100% specificity for all subcomponents. Dice-scores of 0.85 and 0.80 were obtained for ED and NEN, respectively. The MVF procedure increased accuracies across all assessments by 1%–2% (Table 1).

### BraTS2017 Validation Data Set

The network achieved Dice scores of 0.90, 0.80, and 0.78 with Hausdorff distance of 6.5 mm, 8.7 mm, and 5.5 mm for WT, TC,



and ET, respectively, on the BraTS2017 validation data set (Table 2). Sensitivities were 0.90, 0.80, and 0.78 for WT, TC, and ET, respectively, with 100% specificity for all subcomponents.

**BraTS2018 Validation Data Set**

The network achieved Dice scores of 0.90, 0.82, and 0.80, with Hausdorff distance of 6.0 mm, 7.5 mm, and 4.4 mm for WT, TC, and ET, respectively, on the BraTS2018 validation data set (Table 2). Sensitivities were 0.91, 0.81, and 0.81 for WT, TC, and ET, respectively, with 100% specificity for all subcomponents.

**Clinical Validation Data Set**

The network achieved Dice scores of 0.85, 0.80, and 0.77 with Hausdorff distance of 5.74 mm, 5.94 mm, and 4.06 mm for WT, TC, and ET, respectively, on the Oslo clinical data set (Table 3).

Sensitivities were 0.86, 0.79, and 0.77 for WT, TC, and ET, respectively, with 100% specificity for all subcomponents.

**DISCUSSION AND CONCLUSION**

Gliomas are the most common primary brain tumor. Currently, the vast majority of clinical and research efforts to evaluate response to therapy rely on gross geometric measurements. Manual tumor segmentation is a tedious, time-intensive task that requires a human expert. Quantitative evaluations of manual tumor segmentations have revealed considerable disagreement reflected in Dice scores in the range 74%–85% (13). To address these shortcomings, automated machine learning algorithms have been developed to segment gliomas (2, 3, 6).

MRI-based glioma segmentation algorithms represent a method to reduce subjectivity and provide *quantitative* analysis.

**Table 1.** Cross-Validation Results and Mean Dice-Scores (Across Subjects) on 20 Subjects' Held-Out Data Set

	Fold 1	Fold 2	Fold 3	Average
Cross-Validation Results				
Whole Tumor	0.93	0.94	0.90	0.92
Tumor Core	0.89	0.84	0.80	0.84
Enhancing tumor	0.84	0.80	0.77	0.80
Nonenhancing and Necrosis	0.80	0.81	0.80	0.80
Edema	0.85	0.86	0.85	0.85
	<b>Before MVF</b>		<b>After MVF</b>	
Results on 20 Subjects' Held-Out Data Set				
Whole Tumor	0.89		0.90	
Tumor Core	0.82		0.84	
Enhancing Tumor	0.78		0.80	
Nonenhancing and Necrosis	0.79		0.80	
Edema	0.83		0.85	

Accurate, reproducible, and efficient tumor segmentation has the potential to improve glioma management by being able to differentiate active tumor from necrotic tissue and ED. Therefore, significant efforts have been made to facilitate the progress of automated glioma segmentation.

Our algorithm performed similarly to previously published high-performing algorithms in segmenting ET and WT (Table 2) on the BraTS2017 data set and was one of the top 3 performers in segmenting TC. The algorithm was also one of the top 3 performers in segmenting WT and ET on the BraTS 2018 data set (Table 2).

To generalize the network's performance and evaluate its reliability, we also performed 3-fold cross-validation, which showed mean accuracies of 0.92, 0.84, and 0.80 for WT, TC, and ET, respectively. The results of this cross-validation are not comparable to the accuracies reported by the BraTS challenge, as our

cross-validation used one-third of the data for training, whereas most developers use all 285 cases to train their algorithm and cross-validation is not reported.

Our entire pipeline including all the preprocessing steps took ~5 minutes per subject for testing. The Dice scores were slightly reduced when validated on the clinical data set. This decrease in performance was expected owing to practical considerations when using clinical scans. For example, differences in field strength (1.5 T vs 3 T), clinical imaging sequence parameters, and variability in postprocessing may account for the decreased performance. Despite these limitations our deep learning network was able to segment tumors and subcomponents with excellent results without any fine-tuning, and it shows promise for incorporation into clinical workflow.

Variable performance among CNNs can also be owing to differences in the underlying network architecture. The triple network architecture described here has several advantages when compared with multilabel CNNs. Training 3 separate networks for individual binary segmentation tasks is less complex and less computationally challenging than training one network to perform multiclass segmentations. In addition, as all 3 networks are trained separately as binary segmentation problems, misclassification is also highly reduced, thereby reducing overfitting. The dense architecture also reduces false positives, because all feature maps are reused such that every layer in the architecture receives a direct supervision signal (20). The *vanishing gradient problem* is a challenge when using gradient-based learning methods to train neural networks. If the gradient is too small, the neural network weight will not change in value. As more layers with activation functions are added to the network, the gradient can approach zero, making it difficult to train the network. Our algorithm diminished the vanishing gradient problem by using dense networks, which use feature propagation through the dense connection to the subsequent layers. To overcome computational considerations when using a full-size image, our algorithm used a patch-based 3D Dense-Unet (20). An additional unique feature of our network was the procedure for MVF which effectively eliminated false positives. MVF improved network performance across all segmentations. Compared with previously published work on tumor segmentation, our networks used minimal pre- and postprocessing steps.

**Table 2.** Comparison with Best Performers of BraTS 2017 and BraTS 2018 Challenge

Network Type	Whole Tumor	Tumor Core	Enhancing Tumor
Comparison with Best Performers on BraTS 2017 Validation Data			
EMMA (val)	0.901	0.797	0.738
Wang et.al (val) (Cascaded Network)	0.905	0.837	0.785
Dense UNet (ours) (val)	0.907	0.804	0.787
Comparison with Best Performers on BraTS 2018 Validation Data			
NVIDIA (val)	0.910	0.866	0.823
No New-Net (val)	0.908	0.854	0.810
McKinley et. al (val)	0.900	0.853	0.794
Dense Unet (ours) (val)	0.900	0.820	0.800

**Table 3.** Mean Dice-Scores (Across Subjects) on Clinical Data Set

Tumor Type or Subcomponent	Dice Scores
Whole tumor	0.85
Tumor Core	0.80
Enhancing tumor	0.77
Edema	0.80
Necrosis	0.74

## LIMITATIONS

A general limitation of deep learning methods is the need for a large number of subjects to train the network. The BraTS server validation data set does not provide Dice coefficients for “nonenhancing tissue + necrosis,” and edema labels. Even though our network was able to identify these components, the performance for these labels could not be evaluated using the BraTS validation

## ACKNOWLEDGMENTS

This work was supported by NIH/NCI U01CA207091 (AJM, JAM), the Research Council of Norway Grant 191088/F20 (KEE, AB), and the South-Eastern Norway Regional Health Authority Grant 2013069 (KEE).

## REFERENCES

- Shreyas V, Pankajakshan V. A deep learning architecture for brain tumor segmentation in MRI images. In: *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*, Luton; 2017:1–6.
- Pei L, Reza SMS, Li W, Davatzikos C, Iftekharuddin KM. Improved brain tumor segmentation by utilizing tumor growth model in longitudinal brain MRI. *Proc SPIE Int Soc Opt Eng*. 2017.
- Kamnitsas K, Bai W, Ferrante E, McDonagh S, Sinclair M, Pawlowski N, Rajchl M, Lee M, Kainz B, Rueckert D, Glocker B. Ensembles of multiple models and architectures for robust brain tumour segmentation. In: Crimi A, Bakas S, Kuijff H, Menze B, Reyes M, eds. *Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries*. BrainLes 2017. Lecture notes in computer science, Vol. 10670. Springer: Cham; 2018:450–462.
- Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slobboom J, Wiest R, Lanczi L, Gersner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp C, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Trans Med Imaging*. 2015;34:1993–2024.
- Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci Data*. 2017;4:170117.
- Wang G, Li W, Ourselin S, Vercauteren T. Automatic brain tumor segmentation based on cascaded convolutional neural networks with uncertainty estimation. *front comput neurosci*. 2019;13:56.
- Pereira S, Pinto A, Alves V, Silva CA. Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Trans Med Imaging*. 2016;35:1240–1251.
- Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, Rueckert D, Glocker B. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal*. 2017;36:61–78.
- Dvorak PM. Structured prediction with convolutional neural networks for multimodal brain tumor segmentation. In: Menze BH, Reyes M, Farahani K, Kalpathy-Cramer J,

server data set. Memory and computation power limitations remain a consideration in deep learning methods. For instance, when overlapping patches were increased from 75% to 85% for our algorithm, the 3-fold cross-validation results increased to 0.92, 0.84, and 0.80 for WT, TC, and ET, respectively. However, owing to memory limitations, the 85% overlapping patches could not be implemented for training by using all the 265 subjects. This suggests that the network has room for improvement with additional memory and computational power advancements.

## CONCLUSION

A 3D Dense UNet was developed for MRI-based segmentation of gliomas. The algorithm can easily be incorporated into a clinical workflow. Our algorithm outperformed the best performers for segmenting WT and ET.

## Supplemental Materials

Supplemental Data: <https://doi.org/10.18383/j.tom.2019.00026.sup.01>

Conflict of Interest: None reported

Disclosures: No disclosures to report.

- Kown D, eds. *Proceedings of the Multimodal Brain Tumor Segmentation Challenge (MICCAI-BRATS)*. Springer: Munich, Germany; 2015:13–24.
- Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, Pal C, Jodoin P-M, Larochelle H. Brain tumor segmentation with deep neural networks. *Med Image Anal*. 2017;35:18–31.
- Zhuge Y, Krauze AV, Ning H, Cheng JY, Arora BC, Camphausen K, Miller RW. Brain tumor segmentation using holistically nested neural networks in MRI images. *Med Phys*. 2017;44:5234–5243.
- Zikic D, Ioannou Y, Brown M, Criminisi A. Segmentation of brain tumor tissues with convolutional neural networks. In: *Proceedings MICCAI Multimodal Brain Tumor Segmentation Challenge (BrATS)*; 2014:36–39.
- Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci Data*. 2017;4:170117.
- Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imaging*. 2015;15:29.
- Rohlfing T, Zahr NM, Sullivan EV, Pfefferbaum A. The SRI24 multichannel atlas of normal adult human brain structure. *Hum Brain Mapp*. 2009;31:798–819.
- Emblem KE, Pinho MC, Zollner FG, Due-Tonnessen P, Hald JK, Meling TR, Rapalino O, Bjornerud A. A generic support vector machine model for preoperative glioma survival associations. *Radiology*. 2015;275:228–234.
- Emblem KE, Due-Tonnessen P, Hald JK, Bjornerud A, Pinho MC, Scheie D, Schad LR, Meling TR, Zoellner FG. Machine learning in preoperative glioma MRI: survival associations by perfusion-based support vector machine outperforms traditional MRI. *J Magn Reson Imaging*. 2014;40:47–54.
- Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage*. 2011;54:2033–2044.
- Tustison NJ, Cook PA, Klein A, Song G, Das SR, Duda JT, Kandel BM, van Strien N, Stone JR, Gee JC, Avants BB. Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. *Neuroimage*. 2014;99:166–179.
- Jegou S, Drozdal M, Vazquez D, Romero A, Bengio Y. The one hundred layers tiramisu: fully convolutional densenets for semantic segmentation. In: *2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW)*. IEEE: Honolulu, HI; 2017:1175–1183.
- Wegmayr V, Aitharaju S, Buhmann J. Classification of brain MRI with big data and deep 3D convolutional neural networks. In: *SPIE Proceedings, Medical Imaging 2018: Computer-aided Diagnosis*. SPIE Medical Imaging, 2018: Houston, TX; 2018:1057501.

22. Feng X, Yang J, Lipton ZC, Small SA, Provenzano FA. Deep learning on MRI affirms the prominence of the hippocampal formation in Alzheimer's disease classification. *bioRxiv*. 2018;2018.
23. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M, Kudlur M, Levenberg J, Monga R, Moore S, Murray DG, Steiner B, Tucker P, Vasudevan V, Warden P, Wicke M, Yu Y, Zheng X. Tensorflow: a system for large-scale machine learning. In: *12th Symposium on Operating Systems Design and Implementation*. SIGPLAN Not. Savannah, GA; 2016:265–284.
24. Chollet F, Keras. 2015. <https://keras.io>.
25. Kingma DP, Ba JL. Adam: a method for stochastic optimization. In: *3rd International Conference for Learning Representations, San Diego*; 2015.