

# Volumetric Multimodality Neural Network For Brain Tumor Segmentation

Laura Silvana Castillo<sup>\*</sup>, Laura Alexandra Daza<sup>\*</sup>, Luis Carlos Rivera<sup>\*</sup>, and Pablo Arbeláez

Biomedical Engineering Department, Universidad de los Andes, Carrera 1 N° 18A - 12,  
Bogotá Colombia.

<sup>\*</sup>Authors with equal contribution.

## ABSTRACT

Brain lesion segmentation is a challenging biomedical problem. Here we present a convolutional neural network that produces a semantic segmentation of brain tumors, capable of processing volumetric information from multiple MRI modalities at the same time. This results in the ability to learn from small training datasets and highly imbalanced data. We present a new architecture with three parallel contracting pathways that receive inputs in different resolution and then merges their results using three fully connected layers. We tested our method over the 2015 BraTS Challenge dataset, reaching an average dice coefficient of 84%.

**Keywords:** Semantic segmentation, Brain tumor, Deep learning, MRI

## 1. INTRODUCTION

Brain tumors are abnormal formations of mass that limit the space available for the brain inside the skull; as a result, tumors may injure some areas of the brain. These lesions can be classified as Low-Grade Gliomas (LGG) and High-Grade Gliomas (HGG). LGGs are benign, slowly growing tumors that can become life-threatening in the course of disease. HGGs are malignant, fast growing tumors capable of inducing the development of new tumors in different parts of the brain. Without an appropriate treatment, HGGs can be lethal in a few months.<sup>1,2</sup> Because tumors have different shapes, sizes and internal structures, finding those formations in a medical image is a complicated task, highly prone to error.<sup>3</sup> In spite of the fact that the treatment selection is based on diagnosis, nowadays that process is limited, inefficient and observer-dependent. Magnetic Resonance Imaging (MRI) has been used to identify this kind of structures in the brain, and the responsibility to find whether there is an abnormality or not in the exam lies on the neurologist's hands. These physicians determine, based on their own experience, if there is a lesion and what is the best way to proceed. As a consequence, the uncertainty of the patient's outcome is significant.

In the last years, convolutional neural networks (CNN) have shown outstanding results in detection, classification and segmentation tasks, being able to match and sometimes outperform humans. Some of this success is due to the rapid improvement of machine's computational power and to CNNs ability of abstracting features in different hierarchical representations of an image.<sup>4</sup> In this project, we aim at providing an efficient, accurate way of automatically estimating the volume and location of a brain tumor. To do that, we use the BraTS Challenge 2015<sup>3,5</sup> dataset. This open source dataset has MRIs from 220 patients with HGG and 54 patients with LGG tumors; there are four different MRI modalities per patient and annotations made by several specialists. In terms of methodology, we generalize V-Net,<sup>6</sup> an architecture designed to segment prostate on Computed Tomography (CT) images, allowing it to receive multi-modality information. We also looked into DeepMedic,<sup>7</sup> a method created to segment tumors using information from different MRI modalities. During the testing stage, the average dice coefficient was calculated to measure the performance of the methods. The multi-modality version of V-Net reached a score of 50%, while DeepMedic obtained 74%. Based on that, we designed an architecture with multiple input resolutions that reached a dice coefficient of 84% in the test dataset.

---

Further author information: (Send correspondence to Luis Rivera)

Silvana Castillo: E-mail: ls.castillo332@uniandes.edu.co

Laura Daza: E-mail: la.daza10@uniandes.edu.co

Luis Rivera: E-mail: lc.rivera10@uniandes.edu.co

## 2. RELATED WORK

Automatic brain lesion segmentation has been a topic of interest for more than a decade. Early approaches to this problem were based on the detection of abnormalities using healthy-brain atlases and probability models.<sup>8</sup> Then, results were upgraded with the use of deformable registration fields coupled with Markov Random Fields (MRF).<sup>9</sup> Subsequent approaches using machine learning techniques, such as Random Forest,<sup>10,11</sup> improved the results even more, reaching a maximum dice coefficient of 60% in the clinical testing dataset of the 2012 BraTS Challenge.

With the success of deep learning methods, CNNs specialized in semantic segmentation were implemented. These new approaches enhanced the accuracy of the results significantly. Fully convolutional networks (FCN)<sup>12</sup> proved to be an effective way to do pixel-by-pixel classification by obtaining a mean Intersection over Union (IoU) of 67% in the PASCAL-VOC<sup>13</sup> dataset in 2012, where the task was to produce an accurate segmentation of the 20 different categories presented in different natural images. This method offers the advantage of combining coarse and shallow semantic information from images with an arbitrary input size.<sup>12</sup> In 2015, U-Net, a method based on FCN and specialized in the task of segmenting medical images, was developed. U-Net’s architecture has a contracting path to extract local details and an expanding path to locate the object within the whole image.<sup>14</sup> Recently, an expansion of this method to process tridimensional images was presented. V-Net receives a medical volume and performs a binary voxelwise classification.<sup>6</sup> This method demonstrated a remarkable behavior in the PROMISE 2012 Challenge<sup>15</sup> for prostate segmentation on CT images, obtaining an average dice coefficient of 82%. Another highly relevant method to the segmentation of medical images is DeepMedic, a neural network implemented in *Theano*<sup>16</sup> that obtained top ranking performance in the 2015 BraTS Challenge dataset. DeepMedic takes as inputs multimodal 3D patches extracted from MRIs in different resolutions. It analyzes the information using two pathways, one for each resolution used, and combines their results with fully connected layers to generate a semantic segmentation per category.<sup>7</sup> Our approach is based on the multiple resolution paths and the residual connections, this allowed our approach to get different levels of representations of what is a tumor and to train efficiently.

## 3. METHODOLOGY

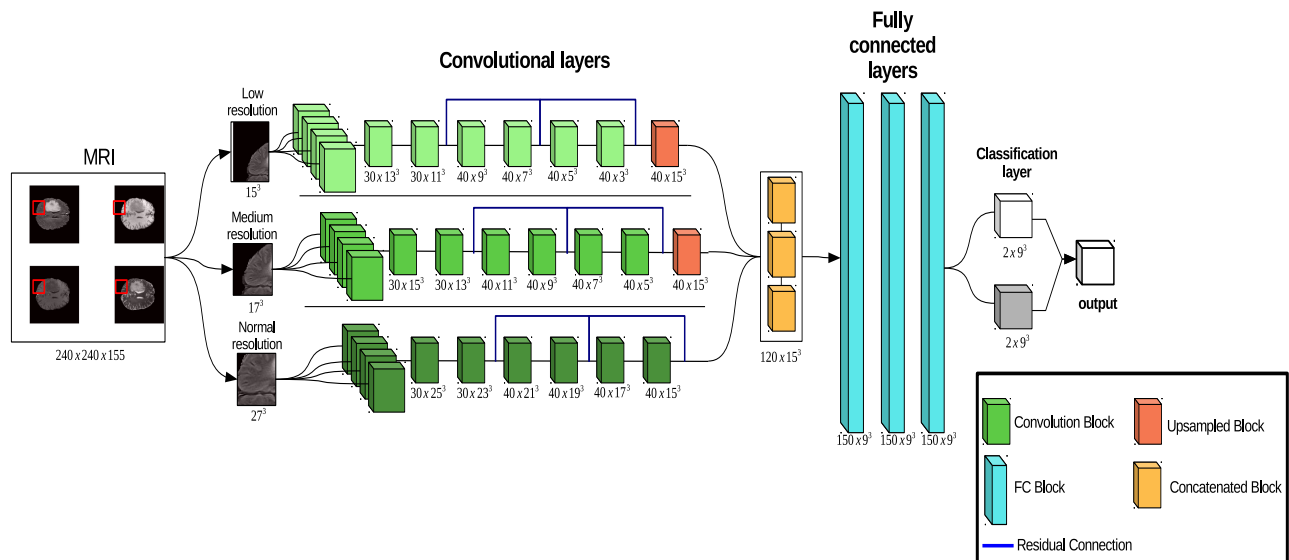


Figure 1: Proposed Architecture. The kernels of the convolutions in the three pathways are  $3^3$  and no padding was made in those operations. The input of the 3 paths are centered in the same voxel, but the medium resolution and low patches are obtained from downsampled versions of the image by factors of 3 and 5, respectively

In this paper, we started analyzing one modality of the MRI and then extended the network to train multimodality volumes on each iteration. We use the 2015 BraTS Challenge dataset for this task without any pre-processing or post-processing, but with data augmentation on-the-fly. This research was developed in the initial stage using the *Caffe*<sup>17</sup> framework and then in *Theano*.<sup>16</sup> To process medical data, we used the SimpleITK package.<sup>18</sup>

### 3.1 Multimodality Volumetric Neural Network

The performance reached by using one MRI modality volume demonstrated that using 3D information allowed the models to obtain robust data for brain tumors segmentation. However, this solution is found to be limited by the selected modality of the image. That dependence is caused by the specific information showed by each of those volumes, causing some to be more useful in the diagnosis of different types of lesions. Based on that, we generalized V-Net<sup>6</sup> to use four MRI modalities at the same time. This architecture has nine layers and is divided into two stages: a convolutional one to obtain fine details of the lesion and a deconvolutional one to extract coarse information. We also used DeepMedic, an architecture designed to learn from different modalities at the same time. DeepMedic has two parallel pathways, with eight convolutional layers each, that receive patches centered at the same voxel but with different resolution in order to obtain fine details and coarse information at the same time.

Given the outstanding results obtained with DeepMedic, we designed a new architecture based on it. Figure 1 shows an overview of our approach. Our architecture has three identical parallel pathways, each one with six convolutional layers and two residual connections, to extract different features on specific resolution levels. All the paths receive patches centered at the same voxel, but extracted from different versions of the image (original and downsampled by factors of three and five). The patches have input sizes of  $27^3$ ,  $21^3$  and  $15^3$  for the normal, medium and low-resolution pathways, respectively. After those last two pathways, an upsample layer is used to make the three outputs of the same size. Finally, the results are concatenated and introduced in the fully connected layers to be combined and then classified. The classification layer is a convolution with a kernel size of  $1^3$  and the final output is made using a softmax classifier layer.

### 3.2 Data

The BraTS challenge (2015) dataset includes 220 patients with HGG and 54 with LGG. Each patient has four MRI modalities (T1, T1 contrast-enhanced, T2, and FLAIR) with a voxel size of  $1mm \times 1mm \times 1mm$ . The ground truth annotations in this dataset contain five different categories, representing the internal structures of the tumor<sup>3,5</sup>:

0. Everything else.
1. Necrosis
2. Edema
3. Non-enhancing tumor
4. Enhancing tumor

Our model is trained to find the whole tumor using binary labels, where 1 represents the 4 internal structures and 0 is everything else.

#### 3.2.1 Training

To train each of the architectures, 137 cases are used, with 110 of those volumes from the HGG set and 27 from the LGG set. To train V-Net, the volumes are downsampled to a size of  $128 \times 128 \times 64$  due to the amount of memory required to process a complete medical volume. To train DeepMedic, patches of size  $25^3$  are extracted randomly, making sure that 50% of them are centered at a voxel labeled as tumor (positive). The same procedure is made to train our architecture, but the patches extracted are of size  $27^3$ . The data is normalized individually by setting the mean to 0 and the variance to 1. Data augmentation is made to avoid overfitting of the model due to the small size of the training dataset, and it is performed on-the-fly to prevent memory issues. The process

is made by reflecting randomly chosen volumes along the sagittal axis. The predictions are made by setting a threshold and classifying every voxel above that value as tumor. To obtain the optimal threshold, different values are tested in all the predictions of the train set, and the one that results in the greater overall dice coefficient is selected.

### 3.2.2 Testing

To test the model, 137 volumes different from the ones used in training are evaluated with the network. The same downscale performed in the training set is made to test V-Net. In the cases of DeepMedic and our architecture, the patches are extracted densely at uniform intervals in the testing set. The selected threshold obtained in the training phase is used to obtain the predictions.

## 3.3 Evaluation Metrics

### 3.3.1 Dice Coefficient

For every model, the Dice-Coefficient (Equation 1) is calculated as performance metric. This measure states the similarity between clinical Ground Truth annotations and the output segmentation of the model. Afterwards, we calculate the average of those results to obtain the overall dice coefficient of the models.

$$DC = \frac{2|A \cap B|}{|A| + |B|} \quad (1)$$

### 3.3.2 Hausdorff Distance

The Hausdorff Distance (Equation 2) is mathematically defined as the maximum distance of a set to the nearest point in the other set,<sup>19</sup> in other words how close are the segmentation and the expected output. This metric works as a complement of the Dice coefficient which gives volumetric similarity while Hausdorff evaluates the boundaries of the segmentation.

$$H(A, B) = \max\{\min\{d(A, B)\}\} \quad (2)$$

### 3.3.3 Sensitivity and Specificity

Are statistical measures used to evaluate the behavior of the predictions and the proportions of True Positives ( $TP$ ), False Negatives ( $FN$ ), False Positives ( $FP$ ) and True Negatives ( $TN$ ). The Sensitivity (Equation 3), also known as True Positive Rate, gives the proportion of true positives predicted correctly. The specificity (Equation 4), also known as True Negative Rate measures, how well the true negatives are predicted.

$$Sensitivity = TPR = \frac{TP}{TP + FN} \quad (3)$$

$$Specificity = TNR = \frac{TN}{TN + FP} \quad (4)$$

## 4. EXPERIMENTAL RESULTS

### 4.1 Volumetric Multimodality using V-net

This neural network was trained with a learning rate of  $1 * 10^{-5}$  and the parameters mentioned before. In Table 1 we present the threshold used, dice coefficient, sensitivity, specificity and Hausdorff distance obtained on the brain tumor segmentation task for each method with a single modality (Flair) and with four modalities. This Table shows that V-Net gets better results using the multimodality approach, however, it is the worst method in all the metrics. In Figure 2 we present some results obtained with this method, where the Ground Truth (GT) column presents the expected output and the second column shows the results obtained with V-Net using four modalities at the same time. In all the predictions the network produced activation points on the correct location, but it reduces the size of the tumors found, which caused a significant decrease in the average dice coefficient and a increase in the Hausdorff distance. However, this architecture predicted fewer false positives than the other two.

## 4.2 Implementation of a multimodality DeepMedic

As mentioned before, we implemented the architecture DeepMedic and trained in the BraTS 2015 dataset with a learning rate of  $5 * 10^{-4}$ . This and the other parameters were selected based on the best values presented by DeepMedic authors.<sup>7</sup> Then modified for those ones which produced a higher overall performance of our network. Table 1 shows that this method also obtains better results with a multimodality approach; It outperforms V-Net improving by 0.24 the dice coefficient, almost doubling the sensitivity and reducing the Hausdorff distance by half. The results displayed in Figure 2 illustrate DeepMedic’s ability to identify the correct location and size of the tumor. However, it also predicts a great amount of false positives regions.

## 4.3 Our Architecture

For our method (Figure 1), we propose a three-pathway architecture, with different resolutions, in order to get information of the location of the tumor and, at the same time, acquire local data that helps to differentiate the parts of the lesions, avoiding false positives. To allow comparison with the previous method, we trained them with the same learning rate used in the implementation of DeepMedic. The different resolutions were selected after trying different ones and selecting those which helped to reduce the high rate of false positives. In Figure 2 we can see the capability of our method to predict the exact area where the patient’s tumor occurs with minimal noise activations in other zones.

Overall, our approach, which uses all modality information, reached a superior result in all the metrics as can be seen in Table 1. The usage of different level resolution and more fully connected layers helped improve the assembly of all feature maps in the trained data, resulting in an average dice coefficient of 84%, which overpassed DeepMedic by 10%, a higher sensitivity and specificity, and a smaller Hausdorff distance. Another strong point of our network is the ability to produce fast segmentations, it takes around fifteen seconds to process the four modalities and produce the new segmentation. In such a critical and time sensitive clinical situation, the speed of these results represents a faster diagnosis and a higher probability of survival for the patient.

Table 1: Dice coefficient, sensitivity, specificity and Hausdorff distance of the baseline (V-Net), the implementation of DeepMedic and Our approach; evaluated over 137 cases of the BraTS 2015 dataset.

	<b>Dice</b>	<b>Sensitivity</b>	<b>Specificity</b>	<b>Hausdorff95</b>
<b>V-Net single modality (Flair)</b>	0,47	0,42	0,983	-
<b>DeepMedic single modality (Flair)</b>	0,61	0,73	0,991	-
<b>Our approach single modality (Flair)</b>	0,66	0,76	0,994	-
<b>V-Net multimodality</b>	0,50	0,47	0,992	113,347
<b>DeepMedic multimodality</b>	0,74	0,79	0,997	54,19758
<b>Our approach multimodality</b>	<b>0,84</b>	<b>0,89</b>	<b>0,998</b>	<b>49,0296</b>

As can be seen in Table 1, the change from using a single modality to using four caused an increase in the average dice result for all the models. However, even if the results with V-Net were not as accurate as the ones with the other two methods, we noticed an interesting behavior: when the V-Net model was tested in the training images, the average dice obtained were 80% and 51% for one and multiple modalities, respectively. This shows an impressive decrease in the over-fitting, resulting in a more general algorithm when adding more modalities.

## 5. CONCLUSIONS

Initially, we wanted to analyze if a multi-modality approach could produce a more robust result than a model trained with just one modality. In order to do this, we tested V-Net and DeepMedic architectures using just the Flair-modality and then four MRI modalities at the same time. We concluded that for all cases multi-modality approach improved the results (Table 1).

Based on V-Net and DeepMedic, we proposed a volumetric multimodality neural network. Our method receives as input 3D patches extracted from the dataset volumes and its architecture consist of three identical parallel pathways, each one with six convolutional layers and two residual connections (to extract features on three specific resolution levels), afterwards there are upsample layers (to make the three outputs of the same size)

and finally, the results are concatenated and introduced in the fully connected layers to be combined and then classified. In this paper, we have presented preliminary results in the 2015 BraTS Challenge dataset reaching an average dice coefficient of 84% overcoming the results of V-Net and DeepMedic (Table 1).

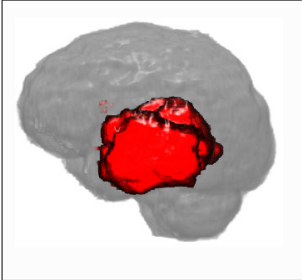
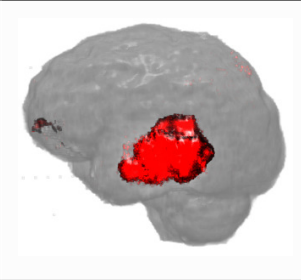
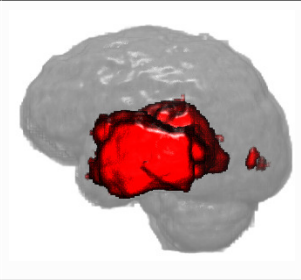
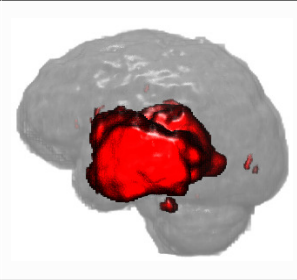
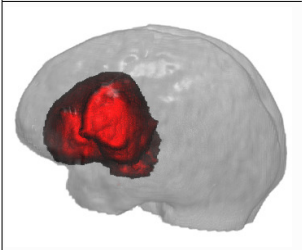
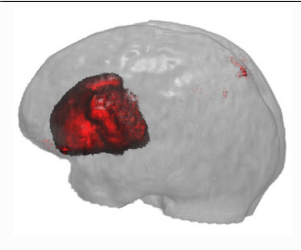
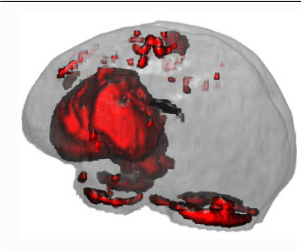
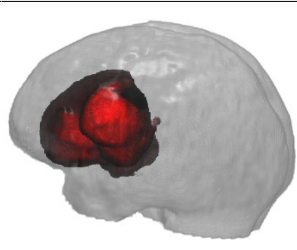
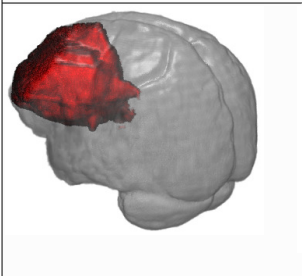
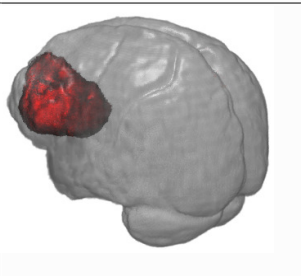
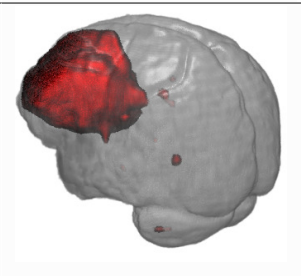
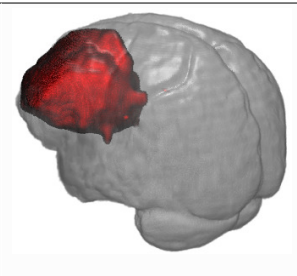
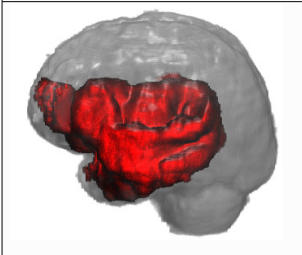
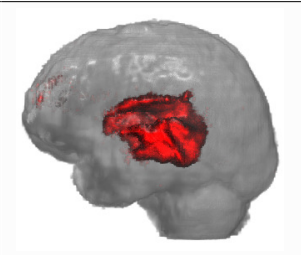
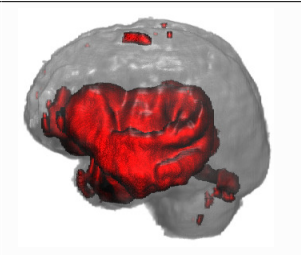
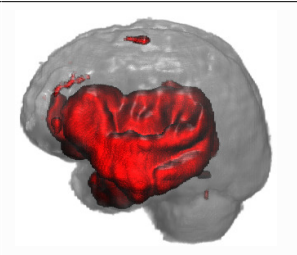
GT	V-Net	DeepMedic	Ours
			
			
			
			

Figure 2: Visual comparison on test dataset. From left to right we present the ground truth (GT), and the predictions of V-Net, DeepMedic and our approach. In all the images red represents a voxel labeled as tumor.

## 6. ACKNOWLEDGEMENTS

The authors gratefully acknowledge NVIDIA Corporation for donating the GPUs used in this project.

## REFERENCES

- [1] “kinderkrebsinfo.de brain tumours - tumours of the central nervous system.” [https://www.kinderkrebsinfo.de/diseases/brain\\_tumours/index\\_eng.html](https://www.kinderkrebsinfo.de/diseases/brain_tumours/index_eng.html). Accessed: 2017-08-24.

- [2] “The Royal Marsden NHS Foundation Trust glioma.” <https://www.royalmarsden.nhs.uk/your-care/cancer-types/paediatric-cancers/glioma>. Accessed: 2017-08-24.
- [3] Menze, B., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.-A., Arbel, T., Avants, B., Ayache, N., Buendia, P., Collins, L., Cordier, N., Corso, J., Criminisi, A., Das, T., Delingette, H., Demiralp, C., Durst, C., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K., Jena, R., John, N., Konukoglu, E., Lashkari, D., Antonio Mariz, J., Meier, R., Pereira, S., Precup, D., Price, S. J., Riklin-Raviv, T., Reza, S., Ryan, M., Schwartz, L., Shin, H.-C., Shotton, J., Silva, C., Sousa, N., Subbanna, N., Szekely, G., Taylor, T., Thomas, O., Tustison, N., Unal, G., Vasseur, F., Wintermark, M., Hye Ye, D., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., and Van Leemput, K., “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS),” *IEEE Transactions on Medical Imaging* **34**(10), 1993,2024 (2015).
- [4] Zeiler, M. D. and Fergus, R., “Visualizing and understanding convolutional networks,” *ECCV* , 818–833 (2014).
- [5] Kistler, M., Bonaretti, S., Pfahrer, M., Niklaus, R., and Büchler, P., “The virtual skeleton database: An open access repository for biomedical research and collaboration,” *J Med Internet Res* **15**, e245 (Nov 2013).
- [6] Milletari, F., Navab, N., and Ahmadi, S., “V-net: Fully convolutional neural networks for volumetric medical image segmentation,” *CoRR abs/1606.04797* (2016).
- [7] Kamnitsas, K., Ferrante, E., Parisot, S., Ledig, C., Nori, A., Criminisi, A., Rueckert, D., and Glocker, B., “Deepmedic on brain tumor segmentation,” *Athens, Greece Proc. BRATS-MICCAI* (2016).
- [8] Prastawa, M., Bullitt, E., Ho, S., and Gerig, G., “A brain tumor segmentation framework based on outlier detection,” *Elsevier* **8**(3), 275–283 (2004).
- [9] Parisot, S., Duffau, H., Chemouny, S., and Paragios, N., “Joint tumor segmentation and dense deformable registration of brain mr images,” *Springer* , 651–658 (2012).
- [10] Bauer, S., Fejes, T., Slotboom, J., Wiest, R., Nolte, L.-P., and Reyes, M., “Segmentation of brain tumor images based on integrated hierarchical classification and regularization,” *Proceedings MICCAI-BRATS* (2012).
- [11] Zikic, D., Glocker, B., Konukoglu, E., Shotton, J., Criminisi, A., Ye, D. H., Demiralp, C., Thomas, O. M., Das, T., Jena, R., and Price, S. J., “Context-sensitive classification forests for segmentation of brain tumor tissues,” *Proceedings MICCAI-BRATS* (2012).
- [12] Shelhamer, E., Long, J., and Darrell, T., “Fully convolutional networks for semantic segmentation,” *CoRR abs/1605.06211* (2016).
- [13] Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A., “The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results.” <http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html>.
- [14] Ronneberger, O., Fischer, P., and Brox, T., “U-net: Convolutional networks for biomedical image segmentation,” *CoRR abs/1505.04597* (2015).
- [15] Litjens, G., Toth, R., van de Ven, W., Hoeks, C., Kerkstra, S., van Ginneken, B., Vincent, G., Guillard, G., Birbeck, N., Zhang, J., et al., “Evaluation of prostate segmentation algorithms for mri: the promise12 challenge,” *Medical image analysis* **18**(2), 359–373 (2014).
- [16] Al-Rfou, R., Alain, G., Almahairi, A., Angermüller, C., Bahdanau, D., Ballas, N., Bastien, F., Bayer, J., Belikov, A., Belopolsky, A., Bengio, Y., Bergeron, A., Bergstra, J., Bisson, V., Snyder, J. B., Bouchard, N., Boulanger-Lewandowski, N., Bouthillier, X., de Brébisson, A., Breuleux, O., Carrier, P. L., Cho, K., Chorowski, J., Christiano, P., Cooijmans, T., Côté, M., Côté, M., Courville, A. C., Dauphin, Y. N., Delalleau, O., Demouth, J., Desjardins, G., Dieleman, S., Dinh, L., Ducoffe, M., Dumoulin, V., Kahou, S. E., Erhan, D., Fan, Z., Firat, O., Germain, M., Glorot, X., Goodfellow, I. J., Graham, M., Gülçehre, Ç., Hamel, P., Harlouchet, I., Heng, J., Hidasi, B., Honari, S., Jain, A., Jean, S., Jia, K., Korobov, M., Kulkarni, V., Lamb, A., Lamblin, P., Larsen, E., Laurent, C., Lee, S., Lefrançois, S., Lemieux, S., Léonard, N., Lin, Z., Livezey, J. A., Lorenz, C., Lowin, J., Ma, Q., Manzagol, P., Mastropietro, O., McGibbon, R., Memisevic, R., van Merriënboer, B., Michalski, V., Mirza, M., Orlandi, A., Pal, C. J., Pascanu, R., Pezeshki, M., Raffel, C., Renshaw, D., Rocklin, M., Romero, A., Roth, M., Sadowski, P., Salvatier, J., Savard, F., Schlüter, J.,

Schulman, J., Schwartz, G., Serban, I. V., Serdyuk, D., Shabanian, S., Simon, É., Spieckermann, S., Subramanyam, S. R., Sygnowski, J., Tanguay, J., van Tulder, G., Turian, J. P., Urban, S., Vincent, P., Visin, F., de Vries, H., Warde-Farley, D., Webb, D. J., Willson, M., Xu, K., Xue, L., Yao, L., Zhang, S., and Zhang, Y., “Theano: A python framework for fast computation of mathematical expressions,” *CoRR* **abs/1605.02688** (2016).

- [17] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., and Darrell, T., “Caffe: Convolutional architecture for fast feature embedding,” in [*Proceedings of the 22nd ACM international conference on Multimedia*], 675–678, ACM (2014).
- [18] Lowekamp, B. C., Chen, D. T., Ibáñez, L., and Blezek, D., “The design of simpleitk,” *Frontiers in neuroinformatics* **7**, 45 (2013).
- [19] Rote, G., “Computing the minimum hausdorff distance between two point sets on a line under translation,” *Information Processing Letters* **38**(3), 123–127 (1991).