

Fully Convolutional Deep Residual Neural Networks for Brain Tumor Segmentation

Peter D. Chang^(✉)

Department of Radiology, Columbia University Medical Center,
New York City, NY, USA
pdc9002@nyp.org

Abstract. In this paper, a fully convolutional residual neural network (FCR-NN) based on linear identity mappings is implemented for medical image segmentation, employed here in the setting of brain tumors. Inspired by deep residual networks which won the ImageNet ILSVRC 2015 classification challenge, the FCR-NN combines optimization gains from residual identity mappings with a fully convolutional architecture for image segmentation that efficiently accounts for both low- and high-level image features. After training two separate networks, one for the task of whole tumor segmentation and a second for tissue sub-region segmentation, the serial FCR-NN architecture exceeds state-of-the-art with complete tumor, core tumor and enhancing tumor validation Dice scores of 0.87, 0.81 and 0.72 respectively. Despite each FCR-NN comprising a complex 22 layer architecture, the fully convolutional design allows for complete segmentation of a tumor volume within 2 s.

1 Introduction

In recent years, convolutional neural networks (CNN) have become the technique of choice for state-of-the-art implementations in various image classification tasks [1–5]. In this study a novel fully convolutional residual neural network (FCR-NN) architecture is proposed for medical image segmentation synthesizing several recently described techniques including: (1) deep residual learning; (2) multi-resolution recursive topology; and (3) fully convolutional neural network architecture.

1.1 Deep Residual Learning

Deep residual learning [4] posits that any traditional neural network layer may be reformulated as a residual function by means of a linear transformation identity mapping with respect to the layer input:

$$x_L = Wx_l + f(x_l) \tag{1}$$

Here the output layer, x_L , is the sum of x_l , the input layer, and $f(x_l)$, a residual change from the original input represented by an arbitrary sequence of non-linear

functions. The projection matrix, W , is an optional term to match input and output layer dimensions. By preserving a linear relationship between the input and output layers, residual neural networks are able to stabilize gradients during backpropagation, leading to improved optimization and facilitating greater network depth. Networks based on this residual architecture have achieved top results in the ImageNet ILSVRC 2015 international challenge.

Image Segmentation Adaptation. In this study we hypothesize that the image segmentation problem can be similarly reformulated as a residual function, such that the final classification labels can be derived from a simple linear combination between the original input image, x_l , and some arbitrary residual, $f(x_l)$. In addition to the advantages for optimization and network depth described above, this architecture benefits from using the original input image directly in the final classification task. This latter proposition may be counterintuitive at first; in fact a widely held assumption is that the raw image inputs contain too much inherent noise and variation to be useful directly for classification. However unique to medical image segmentation is that after accounting for high-level contextual anatomic cues, final segmentation is heavily influenced by low-level features of the original image, namely the signal intensity of any given voxel. For this reason we hypothesize that a deep residual architecture may be optimally suited for the medical image segmentation task.

1.2 Residual Functions for Multi-resolution Features

It is well-established that for the task of object localization a synthesis of contextual cues at multiple spatial resolutions is required for optimal performance [3, 6, 7]. For neural networks this has been implemented by combining feature maps from various layers, with the convolutional transpose operation used to upsample high-order deep layer activations to match those of more superficial layers [8]. Ronneberger et al. [9] further elaborated on this technique by proposing a symmetric contracting and expanding topology that efficiently combines low- and high-level features.

This study reformulates the latter strategy in the form of residual functions. Specifically, from Eq. (1) above, an explicit definition is made such that the residual function, f_l , represents some arbitrary nonlinear transformation performed on the original image, x_l , that captures its higher-level features and which is subsequently combined with its original lower-level properties. A more precise formulation of the higher-level feature map, x_{l+1} , is given by:

$$x_{l+1} = \sigma_l(x_l) \tag{2}$$

Here $\sigma_l(x_l)$ represents a sequence of convolutions and nonlinear activation functions which act to transform the original image, x_l , to its higher-level feature map, x_{l+1} . As a result, the original function f_l in Eq. (1) which was defined in the previous paragraph as the residual arising from the influence of higher-level features can be written as:

$$f_l(x_l) = x_{l+1} + f_{l+1}(x_{l+1}) = \sigma_l(x_l) + f_{l+1}(\sigma_l(x_l)) \quad (3)$$

Here f_{l+1} is a new residual function which like its predecessor represents an incremental change arising from the influence of even higher-level features from the first feature map x_{l+1} . Expanding Eq. (1) recursively allows us to formulate the neural network for image segmentation as a simple series of residual functions, each term of which represents the additional incremental change of gradually higher-level spatial features:

$$x_L = Wx_l + f_l(x_l) = Wx_l + \sigma_l(x_l) + \sigma_{l+1}(\sigma_l(x_l)) + \dots \quad (4)$$

The formulations as described by Eqs. (3) and (4) are visually demonstrated in Fig. 1.

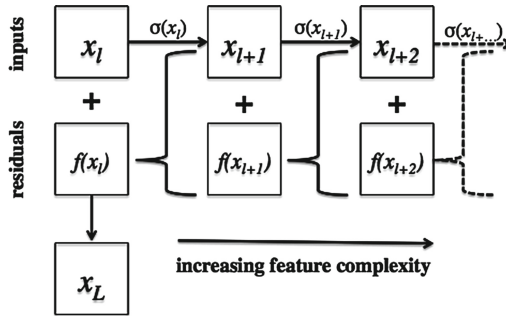


Fig. 1. Recursive series of residual functions for image segmentation.

1.3 Fully Convolutional Neural Networks

Originally introduced by Long et al. [8] fully convolutional neural networks are implemented by a series of upsampling convolutional transpose operators performed on the deepest network layers, resulting in a final dense classification matrix equal in dimension to the original image size for each forward pass. In contradistinction to the typical sliding window CNN approach for image segmentation [10–12], a fully convolutional architecture is highly efficient. In the proposed FCR-NN where an entire axial MR slice is used as the input image, the number of required forward passes for classification of each patient is equal only to the number of slices in the MR volume, a task that can be achieved in less than a few seconds on a GPU optimized workstation.

Several additional less apparent benefits from fully convolutional neural networks include: (1) increased inherent regularization; and (2) influence of per voxel classification from a larger field-of-view. The first effect relating to increased regularization arises from observation that the loss function for each forward/backward pass is derived from contributions at each voxel within the

input image. In this specific example with an input image size of 240×240 voxels, the loss function is thus driven by 240^2 derivatives per slice. Furthermore the ratio of different tissue classes is reflected accurately by the conglomerate of axial slices in each mini-batch, and thus there is no requirement for multiphase training to account for class imbalances as is typically the case for non-fully convolutional designs [10].

Secondly, compared to a sliding scale CNN design whereby classification is limited to a small input patch, a fully convolutional network allows voxel-wise prediction to be influenced by a larger field-of-view proportional to the network depth. This effect is further complemented by the fact that residual networks allow for relatively deep architectures. In the proposed FCR-NN, convolution of the deepest 15×15 feature map with a series of three 3×3 kernels results in an effective receptive field that covers nearly one quarter (7×7) of the feature map.

2 Related Work

Despite significant research over the past several decades, brain tumor segmentation remains a challenging task. Common strategies cited in the literature include: edge-based methods including active contours [13]; region-based methods [14]; classification or clustering methods with constraints based on atlases [15, 16], deformable models [17] or neighborhood regularization [18]; or hybrid generative-discriminative frameworks [19–24]. However these traditional methods are limited to a priori assumptions about a set of rules or features that best model the segmentation mask.

Convolutional neural networks are an emerging technique in computer vision increasingly recognized as the state-of-the-art approach for various image recognition tasks [25]. The power of neural networks arises from its unique capacity for unsupervised feature extraction, independently learning a high-order representation of the data that best approximates any given problem. In recent years, CNN-based algorithms have been adopted for various medical image segmentation tasks [1–5, 9]. In the BRaTS 2015 challenge, the 2nd [26] and 4th [10] ranked submissions were based on CNN architectures.

3 Network Architecture

A fully convolutional residual neural network (FCR-NN) is proposed for medical image segmentation. The network is composed by a series of recursive residual identity mapping blocks, the prototype of which is shown in Fig. 1.

The network comprises a total of 22 layers (18 regular convolutional kernels and 4 upsampling convolutional transpose kernels) and 661,700 parameters. The entire network architecture for a given single FCR-NN is diagrammed in Fig. 2.

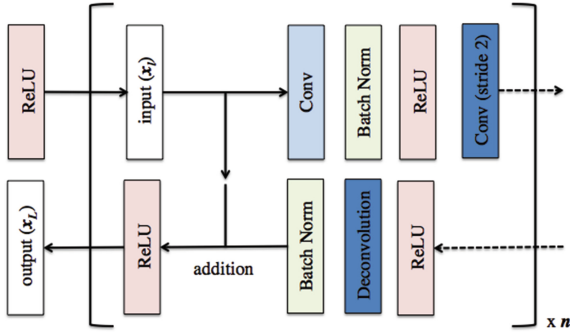


Fig. 2. Residual identity mapping blocks for image segmentation.

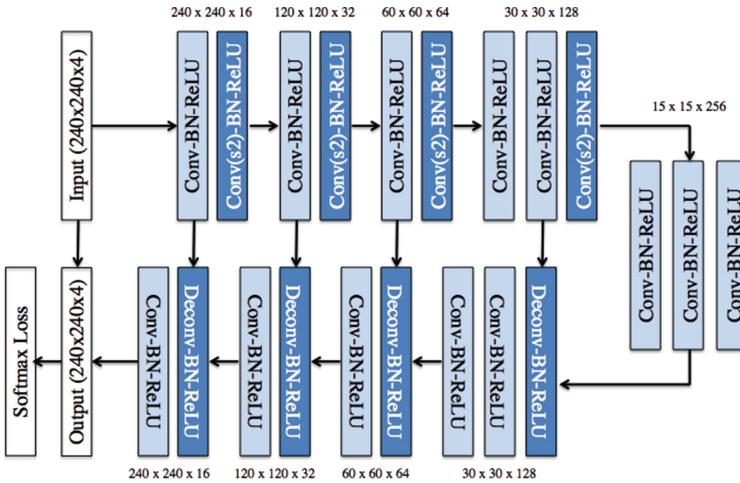


Fig. 3. Fully convolutional residual neural network (FCR-NN) architecture for image segmentation. Abbreviations: Conv = 3×3 convolution; Conv(s2) = 3×3 convolution with stride 2 (downsample); Deconv = 2×2 deconvolution (upsample); BN = batch normalization; ReLU = rectified linear unit

3.1 Fully Convolutional Residual Neural Networks

Convolutional Neural Networks. CNNs are an adaptation of the traditional artificial neural network architecture whereby stacks of 4D convolutional kernels and nonlinear activation functions act to transform a multidimensional input image into progressively higher-order feature representations [27]. The proposed CNN is implemented completely by 3×3 convolutional kernels to prevent overfitting as described by [2]. No pooling layers are used; instead downsampling is implemented simply by means of a 3×3 convolutional kernel with stride length of 2 to decrease the feature maps by 75% in size. All nonlinear functions are modeled by the rectified linear unit (ReLU) [28]. Batch normalization is used between the

convolutional and ReLU layers to limit drift of layer activations during training [29]. In successively deeper layers the number of feature channels gradually increases from 16, 32, 64, 128 and 256, reflecting increasing representational complexity.

Deconvolutions. To upsample each high-level feature map, a deconvolutional or convolutional transpose operator is implemented by means of a 2×2 kernel with a stride length of 2, resulting in an increase in feature map by 75%. At the same time, the number of feature channels is reduced by 50% to match the corresponding activation layer in the mirror image pathway (Figs. 2 and 3).

Residual Connections. As shown in Fig. 2 residual identity mappings are implemented by means of a channel-wise addition operation between an input and its corresponding deconvoluted feature map with the same activation layer size. The addition operation is performed after batch normalization of the deconvoluted output but before the nonlinear ReLU activation as suggested by [4].

3.2 Serial Architecture

The task of brain tumor segmentation as defined in the BRATS challenge can be divided into two corresponding goals. The first task is in differentiating between normal and abnormal brain tissue (whole tumor segmentation). Subsequently, using this result as an input, the second task is differentiate between various brain tumor tissue types. This process is diagrammed in Fig. 4.

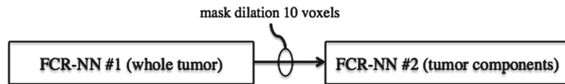


Fig. 4. Implementation of serial fully convolutional residual neural networks (FCR-NN) for the two-part segmentation task.

As shown in the figure, these tasks are learned independently by two separate FCR-NNs. During test time, the segmentation mask generated from the first FCR-NN is dilated by 10 voxels; all voxels outside of this mask are set to 0 and the resulting masked image is used as an input for the second FCR-NN. By separating out these tasks explicitly, the two separate neural networks can be independently optimized for the highest performance for each respective goal.

3.3 Training Details

Image Preprocessing. Each MR volume was normalized simply to a mean signal intensity of 0.5 and standard deviation of $1/6$ such that the range $[0, 1]$ contains three standard deviations in each direction from the mean. No other image preprocessing was necessary.

Data Augmentation. Given the relatively limited training dataset available for the BRATS 2016 challenge, and more generally in the realm of medical imaging where annotated datasets are rare, judicious use of data augmentation is critical for successful CNN implementation. In this study two primary forms of data augmentation are employed. First, a separate cohort of IRB-approved patients at our institution with annotations by a board-certified radiologist were included to increase the breadth of data available for training. Importantly, these included a large number of patients with suboptimal brain extraction and/or significant imaging artifact. In addition, this institutional dataset was predominantly acquired at 3-Tesla MR imaging which, although yields higher sensitivity to pathology, also results in increased image noise and decreased contrast resolution particularly on T2-FLAIR sequences. It is important to point out that upon visual inspection, the test set for the 2016 BRaTS challenge comprised of a disproportionate of 3-Tesla MR imaging relative to the training data available online.

The second type of data augmentation employed by this study involves a number of real-time modifications to the source images at the time of training. Specifically, 50% of all images in a mini-batch were modified randomly by means of: (1) addition across all voxels of a scalar between $[-0.1, 0.1]$; (2) rotation of the image at an angle between 45° to 45° ; (3) addition of a bias field gradient of strength between $[0, 0.4]$ at a random angle between 0° and 360° .

Training Parameters. Training is implemented using standard stochastic gradient descent technique with Nesterov momentum [30]. Parameters are initialized using the heuristic described by He et al. [31]. L2 regularization is implemented to prevent over-fitting of data by limiting the squared magnitude of the kernel weights. To account for training dynamics, the learning rate is annealed and the mini-batch size is increased whenever training loss plateaus. Furthermore a normalized gradient algorithm is employed to allow for locally adaptive learning rates that adjust according to changes in the input signal [32].

Implementation. Software code for this study was written in Matlab (R2016a), and benefited greatly from the MatConvNet toolbox [33]. Experiments were performed on a GPU-optimized workstation with a single NVIDIA GeForce GTX Titan X (12 GB). The combined software and hardware configurations allowed for test time (forward pass only) classification of approximately 155.1 images per second. An entire brain volume with 151 axial slices could be classified in approximately 1.94s (two cumulative forward passes for serial FCR-NNs).

4 Experiments and Results

All 274 cases of gliomas (220 high-grade gliomas; 54 low-grade gliomas) in the BRaTS 2016 training set were included in this study. An additional 150 institutional cases with high-grade gliomas were included to augment the available training data, all acquired on 3T MR scanners. A total of 186 cases in the

BRaTS 2016 database (132/220 high-grade gliomas; 54/54 low-grade gliomas) and 84 institutional cases represented treatment naive (preoperative) tumors. The remaining cases were obtained at various postoperative time points.

Eighty percent of the cases ($n = 339$) were randomly assigned to the training data set, with the remaining 20% of the cases ($n = 85$) used as an independent validation set. All four channels (FLAIR, T1-precontrast, T1-postcontrast, T2) across an entire axial cross-sectional image slice ($240 \times 240 \times 4$ voxels) were used as inputs into the FCR-NN.

Dice scores and Hausdorff distances (mean and range) for the validation set data are reported in Table 1. The results are competitive with historic trends, and finished overall tied for the top two in the BRaTS 2016 challenge.

Table 1. Validation set dice score and Hausdorff distances.

	Complete tumor	Core tumor	Enhancing tumor
Dice	0.89(0.76–0.95)	0.83(0.68–0.91)	0.78(0.41–0.89)
Hausdorff(mm)	8.0(3.0–27)	10.0(4.2–27)	5.9(3.3–21)

Acknowledging limitations in direct comparison with the top four performing algorithms in the BRaTS 2015 challenge due to differences in validation data cited in published results (2013 BRaTS training set in [10, 34]; partial 2015 BRaTS training set in [23]; full 2015 BRaTS training set in [26] which is identical to the 2016 BRaTS training set used in this study), several general patterns can nonetheless be seen. With regards to complete tumor segmentation, the FCR-NN Dice of 0.89 is only marginally better than the previous top four algorithms (0.86–0.88). However given that expert human rater agreement for complete tumor segmentation is reported to be between 0.85–0.91 [35] this metric is likely approaching a plateau for theoretical upper limits of accuracy. By contrast, the much more visually challenging task of separating tumor subcomponents remains to be fully solved. In this case the FCR-NN approach yields a more significant improvement both in Dice score for core tumor (0.83 vs. 0.73–0.79 for previous top performing algorithms) and enhancing tumor (0.78 vs. 0.59–0.73). This is likely in part secondary to a serial architecture which dedicates an entire separate CNN to identification of tumor subcomponents (thus isolating the task from a second CNN that is responsible just for identifying tumor margins).

5 Conclusion

The proposed serial FCR-NN architecture implemented here in the setting of brain tumors is a robust method for medical image segmentation incorporating contextual cues from multiple spatial resolutions. The residual linear transformation identity mappings facilitate optimization and allow for direct contributions from the original input images for final classification. The fully convolutional

architecture results in an efficient classifier which can segment an entire tumor volume in less than 3 s with state-of-the-art accuracy.

Acknowledgments. The author of this paper gratefully acknowledges the support of NVIDIA Corporation with the donation of GeForce GTX Titan X (12 GB) GPU used for this research.

References

1. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks (2012)
2. Simonyan, K., Vedaldi, A., Zisserman, A.: Networks, deep inside convolutional: visualising image classification models and saliency maps. In: ICLR, p. 1 (2014)
3. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 07–12 June, 1–9 Sep 2015 (2015)
4. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. **7**(3), 171–180 (2015). [Arxiv.Org](#)
5. Ciresan, D., Giusti, A.: Deep neural networks segment neuronal membranes in electron microscopy images. In: Advances in Neural Information Processing Systems, pp. 1–9 (2012)
6. Yang, S., Ramanan, D.: Multi-scale recognition with DAG-CNNs (2015)
7. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation (2014)
8. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3431–3440 (2014)
9. 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). doi:[10.1007/978-3-319-24574-4_28](#)
10. 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. In: Proceedings of MICCAI-BRATS (Multimodal Brain Tumor Segmentation Challenge), pp. 29–33 (2015)
11. Kleesiek, J., Urban, G., Hubert, A., Schwarz, D., Maier-hein, K., Bendszus, M., Biller, A.: NeuroImage deep MRI brain extraction: a 3D convolutional neural network for skull stripping. *NeuroImage* **129**, 460–469 (2016)
12. Chen, H., Dou, Q., Yu, L., Heng, P.-A.: VoxResNet: deep voxelwise residual networks for volumetric brain segmentation, pp. 1–9 (2016). [arXiv:1608.05895v1](#)
13. Sachdeva, J., Kumar, V., Gupta, I., Khandelwal, N., Ahuja, C.K.: A novel content-based active contour model for brain tumor segmentation. *Magn. Reson. Imaging* **30**(5), 694–715 (2012)
14. Harati, V., Khayati, R., Farzan, A.: Fully automated tumor segmentation based on improved fuzzy connectedness algorithm in brain MR images. *Comput. Biol. Med.* **41**(7), 483–492 (2011)
15. Prastawa, M., Bullitt, E., Gerig, G.: Simulation of brain tumors in MR images for evaluation of segmentation efficacy. *Med. Image Anal.* **13**(2), 297–311 (2009)

16. Menze, B.H., Leemput, K., Lashkari, D., Weber, M.-A., Ayache, N., Golland, P.: A generative model for brain tumor segmentation in multi-modal images. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6362, pp. 151–159. Springer, Heidelberg (2010). doi:[10.1007/978-3-642-15745-5_19](https://doi.org/10.1007/978-3-642-15745-5_19)
17. Hamamci, A., Kucuk, N., Karaman, K., Engin, K., Unal, G.: Tumor - cut: segmentation of brain tumors on contrast enhanced MR images for radiosurgery applications. *IEEE Trans. Med. Imaging* **31**(3), 790–804 (2012)
18. Zhu, Y., Young, G.S., Xue, Z., Huang, R.Y., You, H., Setayesh, K., Hatabu, H., Cao, F., Wong, S.T.: Semi-automatic segmentation software for quantitative clinical brain glioblastoma evaluation. *Acad. Radiol.* **19**(8), 977–85 (2012)
19. Menze, B.H., Geremia, E., Ayache, N., Szekely, G.: Segmenting glioma in multi-modal images using a generative-discriminative model for brain lesion segmentation. In: Proceedings of MICCAI-BRATS (Multimodal Brain Tumor Segmentation Challenge), p. 7 (2012)
20. Meier, R., Reyes, M., Bauer, S., Slotboom, J., Wiest, R.: A hybrid model for multimodal brain tumor segmentation. In: Proceedings of NCI-MICCAI BRATS (Multimodal Brain Tumor Segmentation Challenge), pp. 31–37 (2013)
21. Bakas, S., Zeng, K., Sotiras, A., Rathore, S., Akbari, H., Gaonkar, B., Rozycki, M., Pati, S., Davatzikos, C.: Segmentation of gliomas in multimodal magnetic resonance imaging volumes based on a hybrid generative - discriminative framework. In: Proceedings of MICCAI-BRATS (Multimodal Brain Tumor Segmentation Challenge), pp. 5–12 (2015)
22. Menze, B.H., Van Leemput, K., Lashkari, D., Riklin-Raviv, T., Geremia, E., Alberts, E., Gruber, P., Wegener, S., Weber, M.-A., Szekely, G., Ayache, N., Golland, P.: A generative probabilistic model and discriminative extensions for brain lesion segmentation with application to tumor and stroke. *IEEE Trans. Med. Imaging* **35**(4), 933–946 (2016)
23. Bakas, S., et al.: GLISTRboost: combining multimodal MRI segmentation, registration, and biophysical tumor growth modeling with gradient boosting machines for glioma segmentation. In: Crimi, A., Menze, B., Maier, O., Reyes, M., Handels, H. (eds.) BrainLes 2015. LNCS, vol. 9556, pp. 144–155. Springer, Cham (2016). doi:[10.1007/978-3-319-30858-6_13](https://doi.org/10.1007/978-3-319-30858-6_13)
24. Zeng, J., See, A.P., Phallen, J., Jackson, C.M., Belcaid, Z., Ruzevick, J., Durham, N., Meyer, C., Harris, T.J., Albesiano, E., Pradilla, G., Ford, E., Wong, J., Hammers, H.-J., Mathios, D., Tyler, B., Brem, H., Tran, P.T., Pardoll, D., Drake, C.G., Lim, M.: Anti-PD-1 blockade and stereotactic radiation produce long-term survival in mice with intracranial gliomas. *Int. J. Rad. Oncol. Biol. Phys.* **86**(2), 343–349 (2013)
25. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436–444 (2015)
26. Pereira, S., Pinto, A., Alves, V., Silva, C.A.: Deep convolutional neural networks for the segmentation of gliomas in multi-sequence MRI. In: Crimi, A., Menze, B., Maier, O., Reyes, M., Handels, H. (eds.) BrainLes 2015. LNCS, vol. 9556, pp. 131–143. Springer, Cham (2016). doi:[10.1007/978-3-319-30858-6_12](https://doi.org/10.1007/978-3-319-30858-6_12)
27. Lecun, Y., Bengio, Y.: Convolutional networks for images, speech, and time-series (1995)
28. Nair, V., Hinton, G.E.: Rectified linear units improve restricted boltzmann machines (2010)
29. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift, pp. 1–11 (2015). [arXiv:1502.03127](https://arxiv.org/abs/1502.03127)

30. Bengio, Y., Boulanger-Lewandowski, N., Pascanu, R.: Advances in optimizing recurrent networks. In: ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, pp. 8624–8628 (2013)
31. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: surpassing human-level performance on imagenet classification (2015)
32. Mandic, D.P.: A generalized normalized gradient descent algorithm. *IEEE Signal Process. Lett.* **11**(2), 115–118 (2004)
33. Vedaldi, A., Lenc, K.: MatConvNet. In: Proceedings of the 23rd ACM International Conference on Multimedia - MM 2015, pp. 689–692 (2015)
34. Agn, M., Puonti, O., Rosenschöld, P.M., Law, I., Leemput, K.: Brain tumor segmentation using a generative model with an RBM prior on tumor shape. In: Crimi, A., Menze, B., Maier, O., Reyes, M., Handels, H. (eds.) *BrainLes 2015*. LNCS, vol. 9556, pp. 168–180. Springer, Cham (2016). doi:[10.1007/978-3-319-30858-6_15](https://doi.org/10.1007/978-3-319-30858-6_15)
35. Menze, B.H., 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.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, Ç., Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharrudin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M.S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Leemput, K.V.: The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* **34**(10), 1993–2024 (2015)