



# A fusion of salient and convolutional features applying healthy templates for MRI brain tumor segmentation

Petra Takács<sup>1,2</sup> · Levente Kovács<sup>1</sup> · Andrea Manno-Kovacs<sup>1,2</sup> 

Received: 2 March 2019 / Revised: 24 February 2020 / Accepted: 15 September 2020 /

Published online: 23 October 2020

© The Author(s) 2020

## Abstract

This paper proposes an improved brain tumor segmentation method based on visual saliency features on MRI image volumes. The proposed method introduces a novel combination of multiple MRI modalities used as pseudo-color channels for highlighting the potential tumors. The novel pseudo-color model incorporates healthy templates generated from the MRI slices without tumors. The constructed healthy templates are also used during the training of neural network models. Based on a saliency map built using the pseudo-color templates, combination models are proposed, fusing the saliency map with convolutional neural networks' prediction maps to improve predictions and to reduce the networks' eventual overfitting which may result in weaker predictions for previously unseen cases. By introducing the combination technique for deep learning techniques and saliency-based, handcrafted feature models, the fusion approach shows good abstraction capabilities and it is able to handle diverse cases that the networks were less trained for. The proposed methods were tested on the BRATS2015 and BRATS2018 databases, and the quantitative results show that hybrid models (including both trained and handcrafted features) can be promising alternatives for reaching higher segmentation performance. Moreover, healthy templates can provide additional information for the training process, enhancing the prediction performance of neural network models.

**Keywords** Visual saliency · Medical image segmentation · Brain tumor detection · Convolutional neural networks

---

✉ Andrea Manno-Kovacs  
andrea.manno-kovacs@sztaki.hu

Petra Takács  
takacs.petra@sztaki.hu

Levente Kovács  
levente.kovacs@sztaki.hu

<sup>1</sup> Institute for Computer Science and Control SZTAKI, Budapest, Hungary

<sup>2</sup> Faculty of Information Technology and Bionics Pázmány Péter Catholic University, Budapest, Hungary

## 1 Introduction

In the last decade, cancer became one of the leading causes of deaths in higher income countries. The earlier the disease is diagnosed, the higher the chance that the patient can be successfully treated. Therefore, quantitative imaging techniques, such as computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET) play a dominant role in early diagnosis. In the last few years, with the significant improvement of these non-invasive techniques, the emphasis has shifted to the efficient processing of the diverse data.

Gliomas are frequent primary brain tumors in adults [9]. Being highly malignant, this type covers a large portion of all malignant brain tumors. In case of patients with such brain tumors, the role of non-invasive imaging techniques is even more important, as repeated tumor biopsies have a high associated risk. Therefore, continuous monitoring using 3D image modalities (CT, MRI) is a widely applied tool. With the improvement of these sensors, 3D data with high spatial resolution is acquired from the brain, and abnormalities can be detected and monitored, which can help in determining the location, size and shape of the tumor, setting up the accurate diagnosis and also in managing the disease and the treatment process simultaneously. Moreover, by applying biologically variable parameters, like longitudinal relaxation time ( $T1$ ), transverse relaxation time ( $T2$ ), proton density ( $PD$ ) or fluid-attenuated inversion recovery ( $FLAIR$ ) and using varying pulse sequences and imaging parameters, different image contrast parameters can be achieved in MRI [17].

To help automatic glioma detection, the Multimodal Brain Tumor Image Segmentation Benchmark (BRATS) [3, 4, 17] was collected then improved and extended multiple times in the past few years.

When categorizing state-of-the-art tumor segmentation algorithms, we can divide them into two broad families [17]. *Generative* models use detailed prior information about the appearance and spatial distribution of multiple different tissues, including anatomical and domain-specific knowledge to build tumor models to be detected and classified. They usually also perform brain tissue segmentation. In [19], tumors were handled as outliers and detects them by applying a brain atlas, followed by a feature-based tumor segmentation, distinguishing tumor and edema regions. The method in [1] is based on a generative model, where tissues are represented by Gaussian mixture models combined with an atlas-based tissue prior. This model was extended with a tumor prior, using convolutional restricted Boltzmann machines. A bag of words driven robust support vector machine classification model is used in [16], to speed up categorization of benign and malignant brain areas.

The methods belonging to this group can handle unseen images efficiently, but they strongly rely on the registration step: test samples should be accurately aligned to spatial priors, which is problematic for example in the presence of large tumor regions [27].

In the other large group, *discriminative* models use annotated training images and directly learn the characteristics of different segmentation labels without any prior domain knowledge. In [6], support vector machine classification and Conditional Random Field (CRF) based hierarchical regularization were combined for multi-level classification of brain tissues. A CRF framework was also applied in [26] for tumor detection and segmentation together with pixel-pairwise affinity and superpixel-level features. [25] introduced a method, which first applies wavelet-based features, then uses an adaptive artificial neural network for classification. To cover intensity and shape variations of tumors, the methods in this group require huge amounts of training data to thoroughly learn tumor features.

Nowadays, deep learning methods are the most popular models of this group, using convolutional neural networks [18, 21]. Different network architectures, such as U-Net [20]

or cascaded anisotropic networks (WT-Net) [8] are applied for training segmentation models using 2D or 3D interpretation. In the last years, these methods were dominating in the tumor segmentation challenges, for example most of the methods on the leaderboard of the BRATS2015 challenge [10, 12, 13] mainly apply convolutional neural networks for tumor segmentation.

However, the disadvantage of these methods is still their strong dependence on the training data, e. g., they cannot handle images with differing imaging protocols from the ones used for acquiring the training data. They also lack the exploitation of spatial priors, therefore sometimes a post-processing step is added to further enhance the performance.

To compensate for the mentioned drawbacks of the different models, one solution might be to use a mixed generative-discriminative model to fuse handcrafted features and learning [22]. Such model was introduced in [2], with an expectation maximization based generative approach as a first step to segment the volume into tumor and healthy tissue labels. Then, the tumor labels were refined using gradient boosting multi-class classification. Finally, a probabilistic Bayesian strategy was employed for finalizing the tumor segmentation.

From a medical point-of-view, the existence of tumors may support diagnosis, therefore these objects may function as the ROI of the image. This motivates to consider tumors as salient regions in the image, and highlight them by applying a visual saliency [14] model. Our proposed algorithm follows this direction and, inspired by [5], constructs a saliency model using handcrafted features. The referred saliency-based detection algorithm [5] is based on a pseudo-coloring scheme using *FLAIR*, *T2* and *T1c* sequences respectively as RGB channels, followed by a bottom-up color and spatial distance calculation to highlight tumor areas as salient regions in the image.

In our previous paper [23] an improvement of this saliency-based algorithm was proposed. We introduced a novel pseudo-color model applying healthy templates for *FLAIR* and *T2* modalities to further highlight tumor regions. Beside the novel color model, different processing steps were added to improve the segmentation performance. We have also proposed a fusion of saliency and convolutional neural networks (U-Net and WT-Net), and the experiments showed that the fused generative-discriminative model is a promising alternative for efficient tumor segmentation.

The most important contributions of this paper are the following:

1. Introducing further improvements regarding the pseudo-color model, switching to an RGB color analysis approach for saliency estimation.
2. Calculating the pseudo-RGB channels as difference images between a specific image patch and a healthy image template built using the healthy slices of the database for *FLAIR*, *T2* and also the *T1c* sequences.
3. The inclusion of healthy slices, i.e., slices lacking malignant areas, in the process, therefore using the complete database for training and highlighting tumor regions as differences from healthy scans at the same time.

The proposed algorithm follows the same workflow as [23], however, skips the RGB to Lab color conversion of the pseudo color image, which was proposed in the original work [5], and instead, the color-based saliency is calculated on the RGB channels. According to our experiments, by applying the pseudo-RGB difference image in the saliency calculation model, more information is exploited, therefore the segmentation performance is higher than with the converted Lab color space.

The proof-of-concept step of the fusion of the proposed saliency map and the prediction map of the trained convolutional neural networks (U-Net and WT-Net) is further analyzed

and an extensive experimental evaluation is performed. Moreover, the idea of the healthy template based pseudo-RGB difference image is also integrated in the retraining process of the traditional U-Net network.

The evaluation process has been performed on the BRATS2015 dataset [17] which includes ground truth data annotated by experts (see a sample in Fig. 1), therefore creating the possibility for quantitative evaluation. By dividing the database into training and testing parts (by a random split of the dataset, see details later), the original and the proposed methods together with the network-based and the proposed fusion models have been evaluated on 28 randomly selected brain volumes (randomly excluded from the training set), including both high grade glioma (HGG) and low grade glioma (LGG) cases. Moreover, the U-Net retraining with the healthy template based pseudo-RGB difference images was performed on BRATS2015 and BRATS2018 as well.

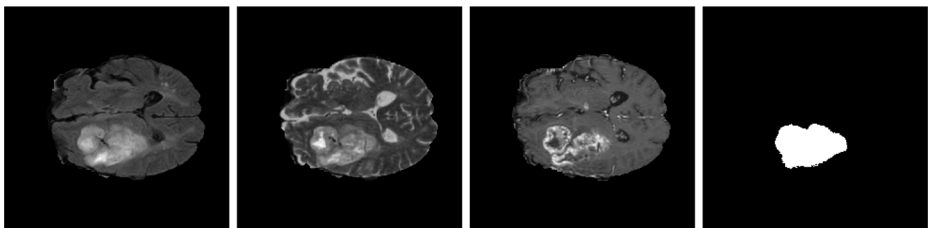
The quantitative results show that the proposed healthy template based pseudo-RGB difference images helped the training and the performance of retrained network models could increase by as much as 8%. The proposed models were compared to the top ranking algorithms of BRATS2015 challenge, and the WT-Net – saliency hybrid model and the retrained U-Net were both able to achieve the same same Dice (DSC) score (0.85) with high Recall values. The experiments show that healthy templates and saliency can be promising additional features which should be further investigated to be integrated in convolutional neural network architectures.

## 2 Visual saliency based tumor segmentation

### 2.1 Pseudo-color model

Inspired by salient object detection algorithms for natural images [11], we construct a color image from the available MRI sequences. Improving the color model of [5] we have also constructed a healthy mean template image for the *FLAIR*, *T2* and *T1c* scans of axial slices in the BRATS2015 database. By analyzing the annotated ground truths, we selected slices without marked tumor regions. For all axial slices, available healthy scans were collected and we constructed the healthy mean templates  $HM_{FLAIR}$ ,  $HM_{T2}$  and  $HM_{T1c}$ . The proposed difference images have the following form:

$$\begin{aligned} D_{FLAIR} &: FLAIR - \alpha * HM_{FLAIR}, \\ D_{T2} &: T2 - \alpha * HM_{T2}, \\ D_{T1c} &: T1c - \alpha * HM_{T1c}, \end{aligned} \quad (1)$$



**Fig. 1** A sample slice from the BRATS2015 data set: Flair, T2, T1c modalities and the ground truth

where we selected  $\alpha = 5/6$  to have a slightly smaller weight on the healthy template and to preserve more the characteristics of the actual scan. We also tested other  $\alpha$  values from  $2/3$  to  $1$  on a smaller test data set (including 20 test volumes from BRATS2015), however the qualitative performance was the highest with the selected value, therefore we defined  $\alpha = 5/6$  empirically. When constructing the difference images, mutual information based registration method [15] is applied.

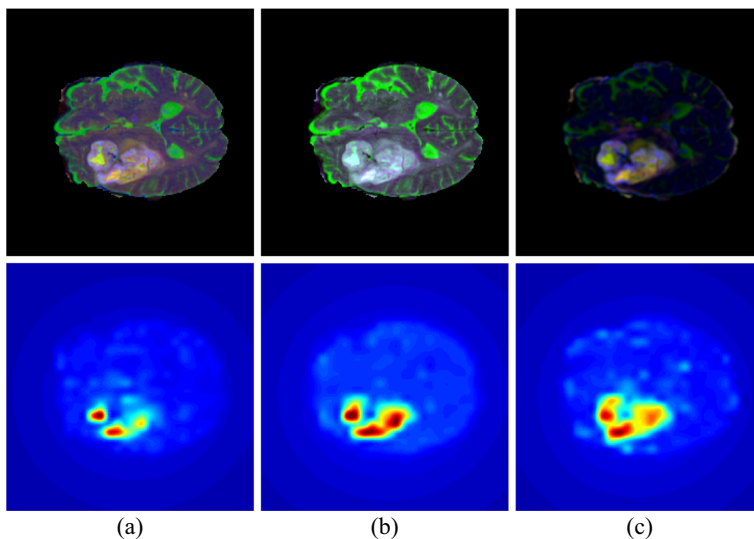
Based on the difference images, the proposed pseudo-RGB model looks as the following:

$$\begin{aligned} \mathbf{R} &: D_{FLAIR}, \\ \mathbf{G} &: D_{FLAIR} * \beta + D_{T2} * \beta, \\ \mathbf{B} &: D_{T1c}, \end{aligned} \quad (2)$$

where  $\beta = 0.5$  is used, to balance between the image characteristics of the different image modalities.

By following the considerations of [24], *FLAIR* and *T2* modalities have both high intensities in peritumoral edema (vasogenic and infiltrative), nonenhancing tumor, white matter injury and gliosis, therefore *FLAIR* based difference image is added to *R* channel. As registration problems may cause the highlighting of areas with cerebrospinal fluid (CSF), especially in *T2* modality, therefore instead of using the  $D_{T2}$  difference image on the *G* channel, a combination of *FLAIR* and *T2* sequences (with equal weights,  $\beta = 0.5$ ) is selected to reduce the misregistration effect (3). To exploit all possible volume information *T1c* based difference image is added to *B* channel.

Instead of transforming the pseudo-RGB image to the CIE  $L^*a^*b^*$  color space, the saliency model is calculated in the RGB space. Figure 2 shows the comparison of the original and proposed pseudo-color models for a more complex case, where the lesion is heterogeneous (see Fig. 1 for the original image modalities and ground truth), therefore the original [5] and the previous [23] models are not able to emphasize the whole area, which



**Fig. 2** Comparison of the pseudo-color models for the sample slice in Fig. 1: **a** the original [5]; **b** the previous [23] and **c** the proposed model. The first row shows the pseudo-color image, the second row shows the computed saliency maps with  $8 \times 8$  block size

results in an inaccurate detection. The proposed model can better highlight the tumor area, even in this complex case.

### 2.2 Saliency map for tumor detection

To build the saliency model, [5] originally advised to apply color difference and spatial difference in a block-based processing system. To achieve this, first the image slice was rescaled to  $256 \times 256$ . Then, the rescaled image was decomposed into non-overlapping blocks with size  $k \times k$ , where  $k = 8$  and  $k = 16$  were used. Therefore, saliency calculation was performed for  $w/k \times w/k$  patches and the  $S_c$  color difference feature for  $R_i$  patch was computed as:

$$S_c(R_i) = \sum_{j, j \neq i} \sqrt{(R_i^{\bar{R}} - R_j^{\bar{R}})^2 + (R_i^{\bar{G}} - R_j^{\bar{G}})^2 + (R_i^{\bar{B}} - R_j^{\bar{B}})^2}, \quad \forall i, j \in \{1, \dots, (w/k \times w/k)\}. \quad (3)$$

The color difference was calculated for each  $RGB$  channels, and  $R_i^{ch}$  marks the mean value for  $ch$  channel, which represents the  $i^{th}$  image patch  $I(R_i)$ :

$$R_i^{ch} = \frac{\sum I(R_i^{ch})}{k \times k}. \quad (4)$$

Further on, the saliency model calculation follows the same procedure which was introduced in our previous work [23]. First, a spatial distance feature is incorporated for saliency calculation:

$$S_{cs}(R_i) = \sum_{j, j \neq i} \frac{1}{1 + d(R_i, R_j)} \times S_c(R_i), \quad (5)$$

where  $d(R_i, R_j)$  is the Euclidean distance of the mean spatial coordinates of  $R_i, R_j$  patches, following the original method [5].

Then, the  $S_{cs}$  saliency map is then scaled back to its original size (denoted by  $\widehat{S}_{cs}$ ), using bilinear interpolation. To make the saliency model scale-invariant to local feature sizes, the  $S_{cs}$  color-spatial saliency is calculated for different block sizes. Using larger block sizes than the magnitude of the tumor regions would cause all detections to fail or induce large errors during the training process by derailing the segmentation steps. On the other hand, too small patches are also not useful and require far too much computation time. Therefore, we applied the same block sizes ( $k = 8, 16$ ) as in [23]. Additionally, we have also tested  $12 \times 12$  instead of  $8 \times 8$  and  $16 \times 16$  block sizes, moreover the performance of 8, 12 and 16 block sizes were also tested together. However, the  $12 \times 12$  block size did not added any extra performance in our experiments and, thus only  $8 \times 8$  and  $16 \times 16$  blocks were used:

$$S_{cs} = \sum_{k=8,16} r_k \times \widehat{S}_{cs,k}, \quad (6)$$

where  $r_k = 0.5$  is applied following the recommendations of [5].

Motivated by the fact that the location, size and shape of the tumor is quite similar in neighboring slices, the final saliency map is calculated as a weighted fusion of the actual ( $S_{cs}^a$ ), previous ( $S_{cs}^p$ ) and next ( $S_{cs}^n$ ) slice’s color-spatial saliency:

$$S = w^p \times S_{cs}^p + w^a \times S_{cs}^a + w^n \times S_{cs}^n, \quad (7)$$

where  $w^p$ ,  $w^a$ ,  $w^n$  denote the weights for the different slices ( $\sum w^i = 1$ ),  $w^a = 0.4$  and  $w^p = w^n = 0.3$  were set, as proposed in [23].

Then a  $25 \times 25$  mean filter is applied on the calculated  $S$  saliency map to get a smoother final estimation.

The saliency map is binarized to have an estimation for the tumor:

$$T_\gamma = \frac{\gamma}{M \times N} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} S(x, y). \quad (8)$$

The original paper [5] proposed  $\gamma = 3.55$ , however the improved pseudo-color model and saliency calculation required a parameter tuning of  $\gamma$ . Thus, different  $\gamma$  values were tested on a smaller data set, including 20 volumes from BRATS2015 for values between 3 and 4.  $\gamma = 3.1$  was selected.

After the saliency map calculation, post-processing steps are introduced in [23] to refine the segmentation result. These steps include a size-based filtering (to eliminate falsely detected areas in healthy slices), an active contour based outline detection for complex tumor shapes and a final drop-out step to eliminate false positive hits, by following tumor candidates throughout neighboring slices and keeping only the detections which appear on the most consecutive slices.

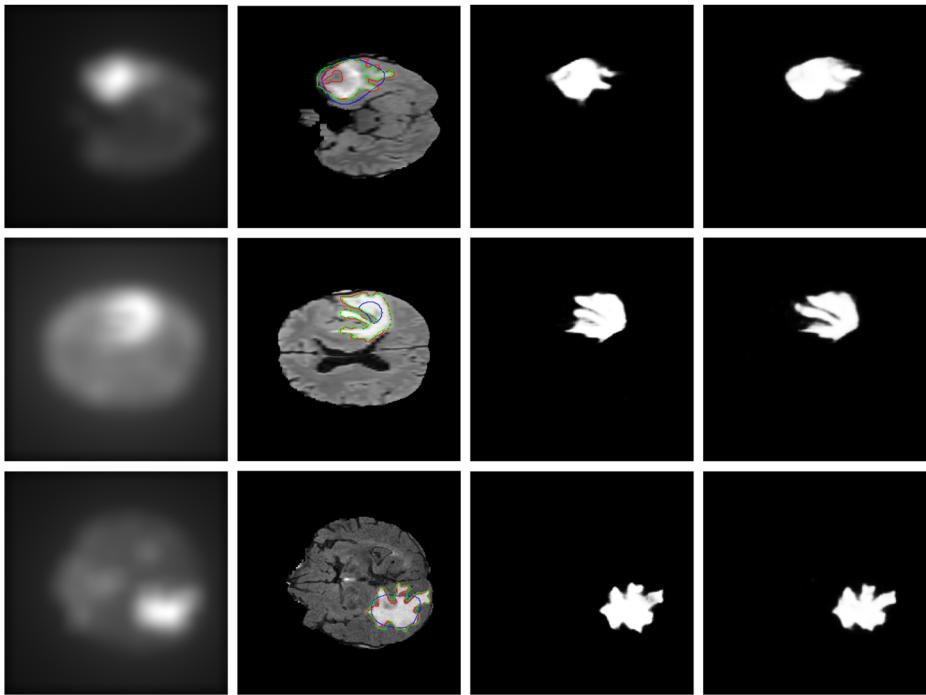
The introduced method is able to locate tumors using the saliency model, then the post-processing steps detect the tumor outlines, even if they have complex shapes. Figure 3 shows a few good examples of the contour detection results, where the saliency-based binarization is shown in blue, the active contour based refinement in red and the ground truth in green. If the applied visual features are strong in the image (i.e. the tumor is differing from its surroundings enough), the saliency model can highlight the tumor successfully and also the active contour can detect the outline because of the intensity difference.

On the other hand, if the tumor cannot be separated from the neighboring tissues with the saliency model (the color-based saliency feature is not highlighting the tumor), then the detection cannot give a precise output. Moreover, the active contour step is also based on image intensity, thus in case of a lesion with less visible outlines, the post-processing cannot give such good detection. Additionally, active contour is an iterative method, and it requires higher computation time, which can be a problem when processing large databases.

As the the second column of Fig. 3 also shows, active contour methods - or possibly any other similar approach - might be capable of producing fairly high quality segmentations in certain situations. Their major drawback comes from the fact that to achieve such high quality results, these methods require extensive parameter tuning and optimization, which is simply not feasible in the case of large amounts of data. This is why our previous [23] and current approach propose a hybrid approach, based on the fusion of a saliency estimation step with neural network predictions for improved, and automatic detection results. In this proposed scheme the training step requires less effort and the trained network will be able to produce predictions faster, more robustly and becomes more scalable.

### 3 Fusion of deep learning prediction maps and handcrafted saliency maps

As it was discussed in the introduction, nowadays neural networks are widely used for brain tumor segmentation. However, adapting deep learning methods to new data can be hard, requiring lengthy retraining, making real world application very challenging. This motivates



**Fig. 3** Tumor contour detection using the Chan-Vese method and by convolutional neural networks: First column is the color-spatial saliency map; Second column is the detection, blue is the thresholded, binary tumor estimation of the color-spatial saliency map, red is the improved result of the active contour step, green is the ground truth tumor outline; Third column is the prediction map of the U-Net [20]; Fourth column is the prediction map of the WT-Net [8]

the idea to fuse a *generative*, handcrafted feature based model and a *discriminative* learning based technique.

Therefore, we have fused our saliency-based model with two, state-of-the-art network architectures, the U-Net [20] and the WT-Net [8]. The U-Net introduces a convolutional network for end-to-end image segmentation, resulting in a segmentation map. The first part of the network is a contractive part, highlighting the image information, while the second part is creating a high-resolution segmentation map (see the third column of Fig. 3). The U-Net was very successful for processing medical image data, used in its original or some minor modified form for segmentation tasks.

In [8] a cascade of CNNs were introduced to segment brain tumor subregions sequentially. The complex segmentation problem is divided into 3 binary segmentation problems: WNet segments the whole tumor, its output is a bounding box, which is used as input of the second network, called as TNet, to segment the tumor core. Finally, its output bounding box is applied as input for ENet to detect enhancing tumor core. As in our case we only concentrate on the whole tumor, and we use the implementation of WNet/TNet, called WT-Net, from the NiftyNet [7] platform, segmentation samples are shown in Fig. 3.

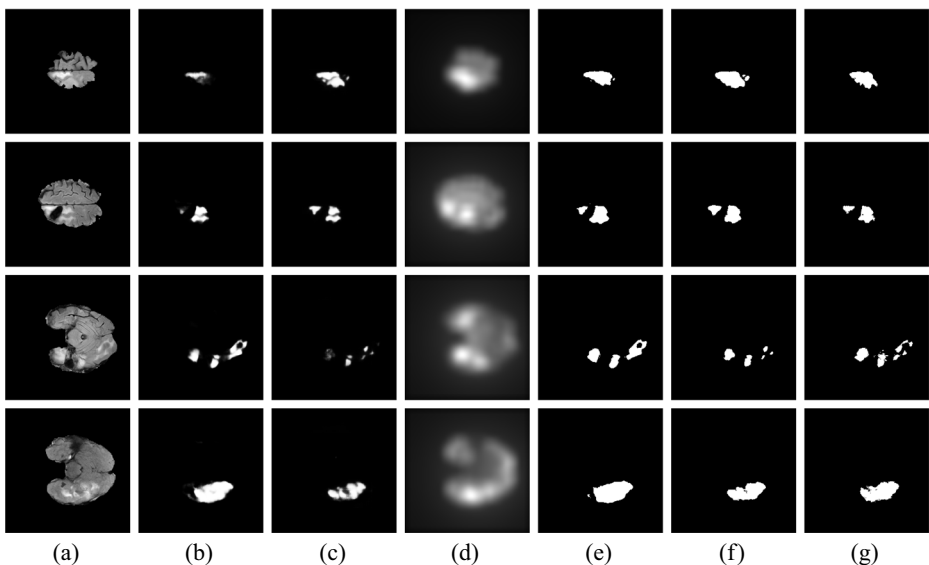


To exploit the benefits of both approaches, the proposed  $S$  saliency map (7) is fused with the prediction map, calculated by the neural network (denoted by  $P_U$  and  $P_{WT}$  for U-Net and WT-Net respectively). As a shallow convolution, the two maps are fused with a weighting function:

$$S_{\{U,WT\}} = \delta \times P_{\{U,WT\}} + (1 - \delta) \times S. \quad (9)$$

Different values were tested for  $\delta$  parameters, which can be seen in Fig. 5, detailed description of the parameter analysis will be given in the experimental section, see Section 4.1. Based on this analysis,  $\delta = 0.775$  was selected for U-Net and  $\delta = 0.7$  was chosen for WT-Net in the quantitative evaluation. According to our experience, while saliency based algorithms have high precision value and lower recall, neural networks behave inversely with higher recall than precision. Moreover, the performance of neural networks with good generalization capabilities can be further improved for unseen, special cases by fusing them with handcrafted features.

The examples in Fig. 4 illustrate the performance of the fusion models: the first two samples are representing how the fusion improves the segmentation result of the U-Net, the third and fourth samples are the improvement for the WT-Net. It should be mentioned, that U-Net and WT-Net produce a probability map as a prediction result, which is binarized in the evaluation process (at 210 threshold value for the  $[0 - 255]$  intensity range). Figure 4b and c show examples of probability maps. The proposed fusion can handle multiple tumor parts (second and third row of Fig. 4) and lesions with heterogeneous regions (first and fourth row of Fig. 4). For more detailed analysis see Section 4.



**Fig. 4** Fusion of neural networks and saliency-based features; **a** FLAIR image slice; **b** result of the U-Net based segmentation; **c** result of the WT-Net based segmentation; **d** calculated saliency map; **e-f** binary segmentation result achieved by the weighted combination (9) with **e**  $\delta = 0.775$  for U-Net and **f**  $\delta = 0.7$  for WT-Net; **g** the binary ground truth for the whole tumor from BRATS2015

## 4 Experimental evaluation

We performed the evaluation on the BRATS2015 dataset [17], which includes altogether 220 HGG volumes and 54 LGG volumes with *T1*, *T1c*, *T2* and *FLAIR* sequences. Each volume has the size of  $240 \times 240 \times 155$  voxels. During the evaluation we used the axial view, i.e., 155 slices with a size of  $240 \times 240$  pixels for each volume. The database includes annotated, pixel-wise ground truth data for all slices.

During evaluation, we used a publicly available implementation<sup>1</sup> of U-Net. All available modalities (T1, T1c, T2 and FLAIR) were used for training with 16-bit slice images and 8-bit ground truth labels. During training, we used a starting learning rate of  $1e - 4$  with Adam optimizer, binary cross entropy loss, with learning rate reduction to a minimum of  $1e - 12$  and early stopping with patience of 6 epochs.

We have also evaluated the NiftyNet [7] implementation of WT-Net. During the training process, the learning rate was set to a constant  $1e - 4$ , the applied loss type was Dice (DSC), and the training was performed for 20000 epochs.

We followed a clear dataset volume separation approach for training-testing dataset generation. For both networks, we followed the same approach for the data selection process for the training and testing/evaluation phases: the dataset was randomly split 90% – 10% for training and testing; then, the training set (the 90% dataset portion just mentioned) was again randomly split 80% – 20% for training and validation during the training process. After the training finished, the separated 10% dataset portion was used for the evaluation (testing) phase.

The randomly partitioned test set included 22 HGG and 6 LGG volumes, including 4340 slices. For every method, the performance was quantitatively evaluated for HGG and LGG volumes separately and also together (marked as HGG+LGG later). For comparison, the same volumes were evaluated for the saliency-based and the fusion models as well (Table 2 and Table 3).

We have calculated different quantitative metrics: Dice score (DSC), Recall (or Sensitivity), Precision (or PPV), and  $F_{\beta}$ :

$$DSC = \frac{2TP}{2TP + FP + FN},$$

$$Recall = \frac{TP}{TP + FN},$$

$$Precision = \frac{TP}{TP + FP},$$

$$F_{\beta} = \frac{(1 + \beta^2)Precision \times Recall}{\beta^2 \times Precision + Recall}, \beta = 0.5 \quad (10)$$

where TP denotes true positives (marked as tumor in the ground truth mask and detected as tumor), FP: false positives (not marked as tumor in the ground truth mask, but detected as tumor), TN: true negatives (not marked as tumor in the ground truth mask and not detected as tumor) and FN: false negatives (marked as tumor in the ground truth mask, but not detected as tumor) respectively. The different values are calculated as the comparison of the ground truth mask and the segmented mask pixelwisely for every slice. In the evaluation

<sup>1</sup><https://github.com/zhixuhao/unet>

the processed image volumes are required to have brain mask (as BRATS databases), the evaluation metrics are only calculated for the brain area, skipping the background.

### 4.1 Weight parameter analysis for fusion models

The first quantitative analysis was performed to select the optimal value for  $\delta$  weight parameter (9) for U-Net and WT-Net in the fusion model. Different  $\delta$  values were tested from 0.6 to 0.9 with 0.025 steps and DSC, Recall, Precision and  $F_\beta$  metrics were measured for HGG+LGG test volumes. Results are shown in Fig. 5, U-Net fusion performance is in the upper image, WT-Net is in the lower one.



Fig. 5 Performance analysis for  $w$  fusion weight parameter for U-Net and WT-Net

We selected the weight parameter based on the highest DSC value, which was  $\delta = 0.775$  for U-Net and  $\delta = 0.7$  for WT-Net. The performance of the fusion models were evaluated with the selected weight parameter throughout the experiments.

## 4.2 Quantitative evaluation of the proposed models

The proposed models were quantitatively evaluated on the BRATS2015 test set, which consisted of randomly selected 22 HGG and 6 LGG volumes. DSC, Recall, Precision and  $F_\beta$  scores were calculated for HGG and LGG volumes separately. To compare the performance of the proposed models, we also evaluated the previous model [23] and the U-Net and WT-Net architectures trained with the traditional image sequences.

The results in Table 1 show some important evaluation results. First, the current proposed model is better in almost all aspects than the previous approach. Secondly, the neural networks alone are performing better than the saliency approach alone. However, the neural networks combined in either a late fusion or a saliency-combined retraining approach are producing improved results.

The data supports the most important point of this paper, that the fusion of saliency information into the neural network based segmentation process is a viable approach and can produce superior results. The advantage of the proposed saliency based approach is its high precision value, which means that the algorithm usually gives smaller, but more precise estimations and the resulting detection is more likely to be inside the real tumor region. On the other hand, U-Net and WT-Net models achieve higher Recall values with a bit lower Precision, meaning that they over-predict the tumor regions, producing areas larger - thus less precise - than the real tumor.

The two behaviours can be beneficially fused, which is well illustrated with the results of the proposed fusion models. The DSC, Precision and  $F_\beta$  scores are significantly increased compared to the original U-Net and WT-Net performance, with slightly decreased Recall values (as described above). The overall performance is still very promising, therefore the

**Table 1** Quantitative results on BRATS2015 dataset for the previously published algorithm [23], the proposed method, U-Net, WT-Net and their weighted combination with the proposed saliency based approach and the U-Net retrained with the healthy template based pseudo-RGB images (highest values are shown in bold)

Algorithm	HGG				LGG			
	DSC	R	P	$F_\beta$	DSC	R	P	$F_\beta$
Previous [23]	0.65	0.64	0.71	0.70	0.60	0.52	<b>0.84</b>	0.58
U-Net [20]	0.80	0.89	0.74	0.76	0.69	0.86	0.58	0.62
WT-Net [7]	0.84	<b>0.95</b>	0.77	0.79	0.75	<b>0.93</b>	0.63	0.67
Proposed saliency model	0.68	0.62	0.82	0.75	0.63	0.62	0.80	0.69
U-Net + saliency ( $\delta = 0.775$ )	0.81	0.87	0.78	0.79	0.73	0.79	0.69	0.70
WT-Net + saliency ( $\delta = 0.775$ )	0.86	0.91	0.84	0.85	<b>0.79</b>	0.85	0.75	0.77
U-Net retrained	<b>0.87</b>	0.87	<b>0.89</b>	<b>0.89</b>	0.78	0.78	0.81	<b>0.80</b>

The randomly selected test sets consist of 22 HGG volumes and 6 LGG volumes. The calculated quantitative metrics are DSC, Recall (R), Precision (P) and  $F_\beta$

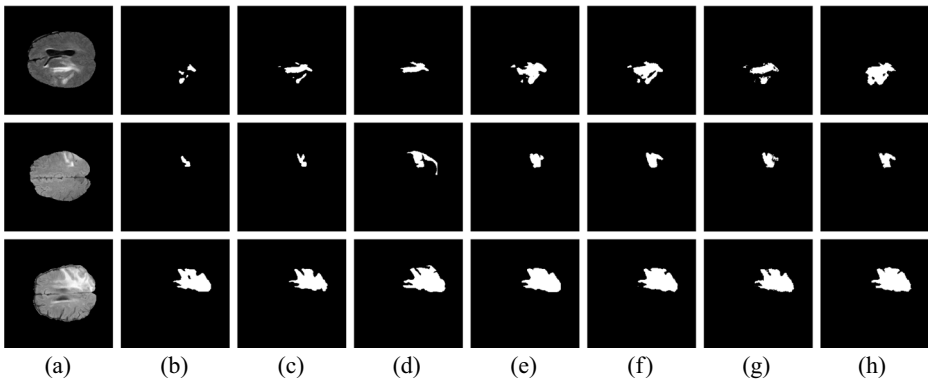
combination of trained networks and handcrafted features (such as saliency) has a great potential for segmentation and it is worth to be further analyzed.

Due to the good performance of the improved pseudo-coloring, the U-Net architecture was retrained with a novel, extended training database. Beside the original MRI volumes with the *FLAIR*, *T2*, *T1* and *T1c* modalities, the improved pseudo-color images based on the healthy templates (3) were constructed for the training slices. The training was performed with the same parameters as in the original case, only the training process was changed by augmenting the original training slices with their associated pseudo-colored saliency maps. The performance of the retrained U-net model increased by 7% for the HGG, 9% for the LGG test volumes, which also confirms that extra information can be extracted by also integrating the healthy slices into the training.

We also performed evaluations on a combined HGG+LGG test set, which provides a much harder setting for evaluations. That is because detection is significantly harder in LGG volumes, thus methods tend to generally perform lower over such data. The compared state-of-the-art methods are the top approaches of the BRATS2015 challenge [10, 12, 13]. Each of them applies convolutional neural networks, [12] and [13] proposed novel three-dimensional convolutional neural networks, in [10], a U-Net based architecture is introduced. In Table 2 we show results compared to state-of-the-art approaches. The point of these results is to show that the proposed saliency fusion approach (last 4 rows) can keep up with the other top performers, all the while providing a robust and versatile approach. The comparison showed that some of the proposed methods perform at the same level as the top approaches of the BRATS2015 challenge. The retrained U-Net architecture has the same DSC score with a Recall that is somewhat lower than the others. The combination of WT-Net and the proposed saliency model, with  $\delta = 0.7$  weight parameter also has the same DSC score, but outperforms the compared 3 approaches in the Recall value. Usually, DSC score supposed to be the most important, which is followed by Recall score. The high Recall value means that the algorithm has high alarm rate, which is favorable in case of malignant region detection in medical data. Therefore, these proposed models are highly competitive compared to the state-of-the-art.

**Table 2** Performance comparison with state-of-the-art methods on the summarized HGG+LGG volumes of BRATS2015 dataset (highest values are shown in bold)

Algorithm	HGG+LGG			
	DSC	Recall	Precision	$F_{\beta}$
Previous [23]	0.64	0.60	0.74	0.67
U-Net [20]	0.77	0.89	0.70	0.73
WT-Net [7]	0.82	<b>0.95</b>	0.74	0.77
Kamnitsas et al. [12]	<b>0.85</b>	0.88	0.85	
Kayalibay et al. [13]	<b>0.85</b>	0.82	<b>0.91</b>	
Isensee et atl. [10]	<b>0.85</b>	0.83	<b>0.91</b>	
Proposed	0.67	0.62	0.81	0.73
U-Net + saliency ( $\delta = 0.775$ )	0.79	0.85	0.76	0.77
U-Net retrained	<b>0.85</b>	0.85	0.87	<b>0.87</b>
WT-Net + saliency ( $\delta = 0.7$ )	<b>0.85</b>	0.90	0.82	0.83



**Fig. 6** Segmentation results for different models on MRI image slices **a** from BRATS2015; **b** U-Net [20]; **c** WT-Net [8]; **d** proposed, saliency-based model; **e** weighted combination of U-Net and the proposed method ( $\delta = 0.775$ ); **f** weighted combination of WT-Net and the proposed method ( $\delta = 0.7$ ); **g** the retrained U-Net with pseudo-color images; **h** the binary ground truth for the whole tumor

Figure 6 includes three examples for the binarized segmentation results of the different models. In the first example, both the saliency-based (Fig. 6d) and the convolutional neural networks (Fig. 6b and c) show undersegmentation. However, by fusing the saliency map and the prediction map, the combined algorithm is able to detect the tumor more accurately. Similarly, by comparing the original U-net result (Fig. 6b) with the retrained (Fig. 6g), the detection is more efficient. In the second row, the saliency-based algorithm oversegments, U-Net and WT-Net undersegment. Again, the fusion models and the retrained U-net model achieves higher performance. The example in the third row is a tumor with a complex shape, for which the fusion models and the trained U-Net are able to enhance the accuracy, compared to the results of the original methods.

To summarize, the proposed fusion models are highly competitive compared to the original methods and state-of-the-art techniques, and saliency is a very promising feature to be combined with neural networks.

### 4.3 Retrained U-Net models with pseudo-color images

The original U-Net architecture was retrained with an extended training image set, including the pseudo-color images. Beside BRATS2015, we have also made experiments on

**Table 3** Performance analysis of traditional and retrained U-Net model on BRATS2015 and BRATS2018 database

Method	BRATS2015				BRATS2018			
	DSC	Recall	Precision	$F_{\beta}$	DSC	Recall	Precision	$F_{\beta}$
U-Net [20]	0.77	0.89	0.70	0.73	0.84	0.80	0.92	0.91
U-Net retrained	0.85	0.85	0.87	0.87	0.84	0.80	0.91	0.89

The test sets include 22 HGG and 6 LGG volumes for BRATS2015 and 21 HGG and 8 LGG volumes for BRATS2018 database

BRATS2018 database. For this data set, 21 HGG and 8 LGG volumes were randomly selected for testing. The training parameters were exactly the same, as for BRATS2015.

Predictions using the models trained with the inclusion of the pseudo-color images improved (BRATS2015) or kept (BRATS2018) the original prediction performance (see Table 3). This suggests that the proposed method has the capability to improve on lesser performing solutions and that a deeper embedding of fused salient features into a network model will further improve these capabilities.

## 5 Conclusion

In this paper, an improved, saliency-based algorithm has been introduced for tumor segmentation in brain MRI image volumes. As an improvement, a new pseudo-color model has been proposed, by building healthy mean image templates for *FLAIR*, *T2* and *T1c* scans to highlight tumors as salient image regions. As a novelty, the proposed pseudo-coloring was also applied for training the U-Net convolutional neural network, exploiting extra information from the healthy slices as well.

The saliency-based model was combined with two different, state-of-the-art convolutional neural networks (U-Net and WT-Net), by introducing a weighting function for the saliency map and the networks prediction maps. The combined map integrates the networks' abstraction and the handcrafted features' ability to also handle special, unseen cases more efficiently. Extensive evaluation was performed to analyze the optimal fusion weights.

Quantitative tests on the BRATS2015 and BRATS2018 dataset and comparison with top state-of-the-art segmentation algorithms confirmed that the proposed fusion models are very promising and can achieve the same performance in DSC together with high Recall values. Saliency is a promising features, which should be further analyzed to be integrated in convolutional neural network architectures.

The U-Net model, retrained with the extended training set, significantly overperformed the original version, which showed that previously unused healthy slices in the training set carry extra information which can help to achieve higher performance.

Our near future plans include building a network architecture that can deeply embed the presented pseudo-color based saliency maps into the network model training itself instead of a post-training fusion process.

**Acknowledgments** This work was partially funded by the Hungarian National Research, Development and Innovation Fund (NKFI) grant nr. KH-126688 and the Hungarian Government, Ministry for National Economy (NGM), grant nr. GINOP-2.2.1-15-2017-00083. This paper was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences and by the Hungarian Government, Ministry of Human Capacities in the New National Excellence Program under grant nr. ÚNKP-18-4-PPKE-132.

**Funding** Open access funding provided by ELKH Institute for Computer Science and Control.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Agn M, Puonti O, Law I, af Rosenschöld P, van Leemput K (2015) Brain tumor segmentation by a generative model with a prior on tumor shape. *Proceeding of the Multimodal Brain Tumor Image Segmentation Challenge*, pp 1–4
2. Bakas S, Zeng K, Sotiras A, Rathore S, Akbari H, Gaonkar B, Rozycki M, Pati S, Davatzikos C (2015) Segmentation of gliomas in multimodal magnetic resonance imaging volumes based on a hybrid generative-discriminative framework. *Proceeding of the Multimodal Brain Tumor Image Segmentation Challenge*, pp 5–12
3. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C (2017) Segmentation labels and radiomic features for the pre-operative scans of the tcga-1gg collection. *The Cancer Imaging Archive*, pp 286
4. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann JB, Farahani K, Davatzikos C (2017) Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Sci Data* 4:170117
5. Banerjee S, Mitra S, Shankar BU, Hayashi Y (2016) A novel GBM saliency detection model using multi-channel MRI. *Plos one* 11(1):e0146388
6. Bauer S, Nolte LP, Reyes M (2011) Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization. In: *International conference on medical image computing and computer-assisted intervention*, pp 354–361
7. Gibson E, Li W, Sudre C, Fidon L, Shakir DI, Wang G, Eaton-Rosen Z, Gray R, Doel T, Hu Y, Whyntie T, Nachev P, Modat M, Barratt DC, Ourselin S, Cardoso MJ, Vercauteren T (2018) NiftyNet: a deep-learning platform for medical imaging. *Comput Methods Program Biomed* 158:113–122
8. Guotai W, Wenqi L, Sebastien O, Tom V (2018) Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In: *Brainlesion: glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer, pp 179–190
9. Holland EC (2001) Progenitor cells and glioma formation. *Curr Opin Neurol* 14(6):683–688
10. Isensee F, Kickingereder P, Wick W, Bendszus M, Maier-Hein KH (2018) Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge. In: *Brainlesion: glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp 287–297
11. Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans Pattern Anal Mach Intell* 20(11):1254–1259
12. Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, Rueckert D, Glocker B (2017) Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Med Image Anal* 36:61–78
13. Kayalibay B, Jensen G, van der Smagt P (2017) CNN-based Segmentation of Medical Imaging Data. [arXiv:1701.03056](https://arxiv.org/abs/1701.03056)
14. Koch C, Ullman S (1987) Shifts in selective visual attention: towards the underlying neural circuitry. In: *Matters of intelligence*. Springer, pp 115–141
15. Mattes D, Haynor DR, Vesselle H, Lewellen TK, Eubank W (2003) Pet-ct image registration in the chest using free-form deformations. *IEEE Trans Med Imaging* 22(1):120–128
16. Mehmood I, Sajjad M, Muhammad K, Shah SIA, Sangaiah AK, Shoaib M, Baik SW (2018) An efficient computerized decision support system for the analysis and 3d visualization of brain tumor. *Multimedia Tools and Applications*, pp 1–26
17. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R et al (2015) The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans Med Imaging* 34(10):1993–2024
18. Pereira S, Pinto A, Alves V, Silva CA (2015) Deep convolutional neural networks for the segmentation of gliomas in multi-sequence MRI. In: *International workshop on brainlesion: glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer, pp 131–143
19. Prastawa M, Bullitt E, Ho S, Gerig G (2004) A brain tumor segmentation framework based on outlier detection. *Med Image Anal* 8(3):275–283
20. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: *International conference on medical image computing and computer-assisted intervention*. Springer, pp 234–241
21. Shaikh M, Anand G, Acharya G, Amrutkar A, Alex V, Krishnamurthi G (2017) Brain tumor segmentation using dense fully convolutional neural network. In: *International MICCAI brainlesion workshop*. Springer, pp 309–319



22. Soltaninejad M, Zhang L, Lambrou T, Yang G, Allinson N, Ye X (2017) MRI brain tumor segmentation using random forests and fully convolutional networks. In: International MICCAI brainlesion workshop, pp 279–283
23. Takács P, Manno-Kovacs A (2018) Mri brain tumor segmentation combining saliency and convolutional network features. In: 2018 International conference on content-based multimedia indexing (CBMI), pp 1–6
24. Villanueva-Meyer JE, Mabray MC, Cha S (2017) Current Clinical Brain Tumor Imaging. *Neurosurgery* 81(3):397–415
25. Virupakshappa AB (2018) Computer-aided diagnosis applied to mri images of brain tumor using cognition based modified level set and optimized ann classifier. *Multimedia Tools and Applications*, pp 1–29
26. Wu W, Chen AY, Zhao L, Corso JJ (2014) Brain tumor detection and segmentation in a CRF (conditional random fields) framework with pixel-pairwise affinity and superpixel-level features. *Int J CARS* 9(2):241–253
27. Zacharaki EI, Shen D, Lee SK, Davatzikos C (2008) ORBIT: a multiresolution framework for deformable registration of brain tumor images. *IEEE Trans Med Imaging* 27(8):1003–1017

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Petra Takaács** is an M.Sc. student at the Pazmany Peter Catholic University, Faculty of Information Technology and Bionics. She received her B.Sc. in computer science from Budapest University of Technology and Economics. She works in a software developer position at the Machine Perception Research Laboratory at the Institute for Computer Science and Control of the Hungarian Academy of Sciences (MTA SZTAKI), Budapest, Hungary. Main research interest: medical applications in information technology.



**Levente Kovács** received M.Sc. in information technology (2002), Ph.D. in image processing and graphics (2007) from the University of Pannonia, Hungary. He is currently senior research fellow with the Machine Perception Research Laboratory at the Institute for Computer Science and Control of the Hungarian Academy of Sciences (MTA SZTAKI), Budapest, Hungary. Managing and research participation in several national and international research projects. Main research areas: image/video feature selection, fusion, indexing, retrieval, object detection, classification, machine learning.



**Dr. Andrea Manno-Kovacs** received her PhD in image processing in 2013 from the Pazmany Peter Catholic University in Budapest, her MSc in computer science from Budapest University of Technology and Economics. She is currently a research fellow with the Machine Perception Research Laboratory, at the Institute for Computer Science and Control of the Hungarian Academy of Sciences (MTA SZTAKI) and a part-time research fellow at Pazmany Peter Catholic University, Faculty of Information Technology and Bionics. She has been the manager of various national and international research projects in the recent years. She is also supervising BSc and MSc students at Pazmany Peter Catholic University. Main interests include image and video processing, feature extraction, saliency models and boundary recognition.