



HAL
open science

Deep Learning-Based Concurrent Brain Registration and Tumor Segmentation

Théo Estienne, Marvin Lerousseau, Maria Vakalopoulou, Emilie Alvarez
Andres, Enzo Battistella, Alexandre Carré, Siddhartha Chandra, Stergios
Christodoulidis, Mihir Sahasrabudhe, Roger Sun, et al.

► **To cite this version:**

Théo Estienne, Marvin Lerousseau, Maria Vakalopoulou, Emilie Alvarez Andres, Enzo Battistella, et al.. Deep Learning-Based Concurrent Brain Registration and Tumor Segmentation. *Frontiers in Computational Neuroscience*, 2020, Multimodal Brain Tumor Segmentation and Beyond, 10.3389/fn-com.2020.00017 . hal-02974826

HAL Id: hal-02974826

<https://hal.science/hal-02974826>

Submitted on 22 Oct 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Deep Learning-Based Concurrent Brain Registration and Tumor Segmentation

Théo Estienne^{1,4,5,6*,‡}, **Marvin Lerousseau**^{1,3,4,5,‡}, **Maria Vakalopoulou**^{1,6},
Emilie Alvarez Andres^{1,4,5}, **Enzo Battistella**^{1,4,5,6}, **Alexandre Carré**^{1,4,5},
Siddhartha Chandra³, **Stergios Christodoulidis**², **Mihir Sahasrabudhe**³, **Roger Sun**^{1,3,4,5},
Charlotte Robert^{1,4,5}, **Hugues Talbot**³, **Nikos Paragios**¹ and **Eric Deutsch**^{1,4,5}

¹ *Gustave Roussy-CentraleSupélec-TheraPanacea Center of Artificial Intelligence in Radiation Therapy and Oncology, Gustave Roussy Cancer Campus, Villejuif, France*

² *ARTORG Center for Biomedical Engineering Research, Bern University, Bern, Switzerland*

³ *CVN, CentraleSupélec, University Paris-Saclay and INRIA Saclay, France*

⁴ *INSERM, U1030, Paris, France*

⁵ *University Paris Sud, UFR de médecine, Paris, France*

⁶ *Laboratoire MICS, CentraleSupélec - University Paris-Saclay, France*

Correspondence*:

Théo Estienne

theo.estienne@centralesupelec.fr

2 ABSTRACT

3 Image registration and segmentation are the two most studied problems in medical image
4 analysis. Deep learning algorithms have recently gained a lot of attention due to their success
5 and state-of-the-art results in variety of problems and communities. In this paper, we propose
6 a novel, efficient, and multi-task algorithm that addresses the problems of image registration
7 and brain tumor segmentation jointly. Our method exploits the dependencies between these
8 tasks through a natural coupling of their interdependencies during inference. In particular, the
9 similarity constraints are relaxed within the tumor regions using an efficient and relatively simple
10 formulation. We evaluated the performance of our formulation both quantitatively and qualitatively
11 for registration and segmentation problems on two publicly available datasets (BraTS 2018 and
12 OASIS 3), reporting competitive results with other recent state-of-the-art methods. Moreover, our
13 proposed framework reports significant amelioration ($p < 0.005$) for the registration performance
14 inside the tumor locations, providing a generic method that does not need any predefined
15 conditions (e.g. absence of abnormalities) about the volumes to be registered.

16

17 **Keywords:** brain tumor segmentation, deformable registration, multi-task networks, deep learning, convolutional neural networks.

[‡] equally contributing authors

1 INTRODUCTION

18 Brain tumors and more specifically gliomas as one of the most frequent types, are across the most
19 dangerous and rapidly growing types of cancer (C. Holland, 2002). In clinical practice, multi-modal
20 magnetic resonance imaging (MRI) is the primary method of screening and diagnosis of gliomas. While
21 gliomas are commonly stratified into Low grade and High grade due to different histology and imaging
22 aspects, prognosis and treatment strategy, radiotherapy is one of the mainstays of treatment (Sepúlveda-
23 Sánchez et al., 2018; Stupp et al., 2014). However, radiotherapy treatment planning relies on tumor manual
24 segmentation by physicians, making the process tedious, time-consuming, and sensitive to bias due to low
25 inter-observer agreement (Wee et al., 2015).

26 In order to overcome these limitations, numerous methods have been proposed recently that try to provide
27 tools and algorithms that will make the process of gliomas segmentation automatic and accurate (Parisot
28 et al., 2016; Zhao et al., 2018). Towards this direction, the multimodal brain tumor segmentation challenge
29 (BraTS) (Bakas et al., 2017b,c,a; Menze et al., 2015) is annually organized, in order to highlight
30 efficient approaches and indicate the way towards this challenging problem. In recent years, most of
31 the approaches that exploit BraTS have been based on deep learning architectures using 3D convolutional
32 neural networks (CNNs) similar to VNet (Milletari et al., 2016). In particular, the best performing
33 approaches use ensembles of deep learning architectures (Kamnitsas et al., 2018; Zhou et al., 2018), with
34 autoencoder regularization (Myronenko, 2018) or they even combine deep learning architectures together
35 with algorithms such as conditional random fields (CRFs) (Chandra et al., 2019). Other top-performing
36 methods at the BraTS 2017 and 2018 used cascaded networks, multi-view and multi-scale approaches
37 (Wang et al., 2017), generic UNet architecture with data augmentation and post processing (Isensee et al.,
38 2018), dilated convolutions and label uncertainty loss (McKinley et al., 2018), and context aggregation and
39 localization pathways (Isensee et al., 2017). A more detailed comparison and presentation of the last years
40 challenges on BraTS is presented and summarized in (Bakas et al., 2018).

41 Image registration is a challenging task for medical image analysis in general and for rapidly evolving
42 brain tumors in particular, where longitudinal assessment is critical. Image registration seeks to determine
43 a transformation that will map two volumes (source and reference) to the same coordinate system. In
44 practice, we seek a volume mapping function that changes the coordinate system of the source volume
45 into the coordinate system of the reference volume. Among the different types of methods employed in
46 medical applications, deformable or elastic registration is the most commonly used (Sotiras et al., 2013).
47 Linear methods are an alternative but in that case a linear global transformation is sought for the entire
48 volume. Deformable registration has been addressed with a variety of methods, including for example
49 surface matching (Robinson et al., 2018; Postelnicu et al., 2009) or graph based approaches (Glocker et al.,
50 2009). These methods have been extended to address co-registration of multiple volumes (Ou et al., 2011).
51 Moreover, some of the most popular methods traditionally used for the accurate deformable registration
52 include (Klein et al., 2009; Avants et al., 2008; Shi et al., 2013). Recently a variety of deep learning
53 based methods have been proposed, reducing significantly the computational time but maintaining the
54 accuracy and robustness of the registration (Dalca et al., 2018; Christodoulidis et al., 2018). In particular,
55 the authors in (Dalca et al., 2018) presented a deep learning framework trained for atlas-based registration
56 of brain MR images, while in (Christodoulidis et al., 2018) the authors present a scheme for a concurrent
57 linear and deformable registration of lung MR images. However, when it comes to anatomies that contain
58 abnormalities such as tumoral areas, these methods fail to register the volumes at certain locations, due to
59 lack of similarity between the volumes. This most of the times ends to complete distortion of the tumor
60 area of the deformed image.

61 To overcome this problem, in this paper, we propose a dual deep learning based architecture that addresses
62 registration and tumor segmentation simultaneously, relaxing the registration constraints inside the predicted
63 tumor areas, providing displacements and segmentation maps at the same time. Our framework bears
64 concept similarities with the work presented in (Parisot et al., 2012) where a Markov Random Field (MRF)
65 framework has been proposed to address both of tumor segmentation and image registration jointly. Their
66 method required approximately 6 minutes for the registration of one pair and the segmentation of one
67 class tumor region was performed with handcrafted features and classical machine learning techniques
68 using only one MRI modality. Moreover, there are methods in the literature that try to address the problem
69 of registration of brain tumor MRI by registering on atlases or MRIs without tumoral regions (Gooya
70 et al., 2010, 2012). Here, we introduce a highly scalable, modular, generic and precise 3D-CNN for
71 both registration and segmentation tasks and provide a computationally efficient and accurate method for
72 registering any arbitrary subject involving possible abnormalities. To the best of our knowledge this is the
73 first time that a joint deep learning-based architecture is presented, showing very promising results in two
74 publicly available datasets for brain MRI. The proposed framework provides a very powerful formulation
75 introducing the means to elucidate clinical, or functional trends in the anatomy or physiology of the brain
76 due to the registration part. Moreover, it enables the modeling and the detection of brain tumor areas due to
77 the synergy with the segmentation part.

2 MATERIALS AND METHODS

78 Consider a pair of medical volumes from two different patients —a source S , and a reference R together
79 with their annotations for the tumor areas (S_{seg} and R_{seg}). The framework consists of a bi-cephalic structure
80 with shared parameters, depicted in Figure 1. During training the network uses as input a source S and
81 a reference R volumes and outputs their brain tumor segmentation masks \hat{S}_{seg} and \hat{R}_{seg} and the optimal
82 elastic transformation G which will project or map the source volume to the reference volume. The goal of
83 the registration part is to find the optimal transformation to transform the source (S) to the reference (R)
84 volume. In this section, we present the details for each of the blocks as well as our final formulation for the
85 optimization.

86 2.1 Shared encoder

87 One of the main differences of the proposed formulation with other registration approaches in the literature
88 is the way that the source and reference volumes are combined. In particular, instead of concatenating
89 the two initial volumes, these volumes are independently forwarded in a unique encoder, yielding two
90 sets of features maps (called *latent codes*) C_{source} and $C_{reference}$ for the source and the reference volumes
91 respectively. These two codes are then independently forwarded into the segmentation decoder, providing
92 the predicted segmentation maps S_{seg} and R_{seg} . Simultaneously, the two codes are merged before being
93 forwarded in the registration decoder — this operation is depicted in the "Merge" block in Figure 1.
94 The motivation behind adopting this strategy is based on forcing the encoder to extract meaningful
95 representations from individual volumes instead of a pair of volumes. This is equivalent to asking the
96 encoder discovering a template, "deformation-free" space for all volumes, and encoding each volume
97 against this space (Shu et al., 2018), instead of decoding the deformation grid between every possible pair
98 of volumes. Besides, from the segmentation point of view, there are no relationship between the tumor
99 maps of the source volume and the reference volume, so the codes to be forwarded into the segmentation
100 decoder should not depend on each other.

101 We tested two merging operators, namely concatenation and subtraction. Both source and reference
 102 images are $4D$ volumes whose first dimension corresponds to the 4 different MRI modalities that are used
 103 per subject. After the forward to the encoder, the codes C_{source} and $C_{\text{reference}}$ are also $4D$ volumes with
 104 the first dimension corresponding to n_f , which is the number of convolutional filters of the last block
 105 of the encoder. Before C_{source} and $C_{\text{reference}}$ are inserted into the registration decoder, they are merged,
 106 outputting one $4D$ volume of size $2 \times n_f$ in the case of the concatenation, and of size n_f for the elementwise
 107 subtraction operator, both leaving the rest of the dimensions unchanged. In particular, the subtraction
 108 presents the following natural properties for every coding image C_I :

- 109 • $\forall C_I \in \mathbb{R}^4 : \text{Merge}(C_I, C_I) = 0$
- 110 • $\forall C_I, C_J \in \mathbb{R}^4 \times \mathbb{R}^4 : \text{Merge}(C_I, C_J) = -\text{Merge}(C_J, C_I)$

111 2.2 Brain tumor segmentation decoder

112 Inspired by the latest advances reported on the BraTS 2018 dataset, we adopt a powerful autoencoder
 113 architecture. The segmentation and registration decoders share the same encoder (Section 2.1) for feature
 114 extraction and they provide brain tumor segmentation masks (\hat{S}_{seg} and \hat{R}_{seg}) for the source and the
 115 reference images. These masks refer to valuable information about the regions that cannot be registered
 116 properly as there is no corresponding anatomical information on the pair. This information is integrated
 117 into the optimisation of the registration component, relaxing the similarity constraints and preserving to a
 118 certain extent the geometric properties of the tumor.

119 Variety of loss functions have been proposed in the literature for the semantic segmentation of 3D medical
 120 volumes. In this paper, we performed all our experiments using weighted categorical cross-entropy loss
 121 and optimising 3 different segmentation classes for the tumor area as provided by the BraTS dataset. In
 122 particular,

$$\mathcal{L}_{\text{seg}} = CE(S_{\text{seg}}, \hat{S}_{\text{seg}}) + CE(R_{\text{seg}}, \hat{R}_{\text{seg}}) \quad (1)$$

123 where CE denotes the weighted cross entropy loss. The cross entropy is calculated for both the source
 124 and reference images and the overall segmentation loss is the sum of the two. Here we should note that
 125 different segmentation losses can be applicable as for example the dice coefficient (Sudre et al., 2017),
 126 focal loss (Lin et al., 2017), e.t.c.

127 2.3 Elastic registration decoder

128 In this paper, the registration strategy is based on the one presented in (Christodoulidis et al., 2018), with
 129 the main component being the 3D spatial transformer. A spatial transformer deforms (or warps) a given
 130 image S with a deformation grid G . It can be represented by the operation,

$$D = \mathcal{W}(S, G),$$

131 where $\mathcal{W}(\cdot, G)$ indicates a sampling operation \mathcal{W} under the deformation G and D the deformed image.
 132 The deformation is hence fed to the transformer layer as sampling coordinates for a backward trilinear
 133 interpolation sampling, adapting a strategy similar to (Shu et al., 2018). The sampling process is then
 134 described by

$$D(\vec{p}) = \mathcal{W}(S, G)(\vec{p}) = \sum_{\vec{q}} S(\vec{q}) \prod_d \max(0, 1 - |[G(\vec{p})]_d - \vec{q}_d|),$$

135 where \vec{p} and \vec{q} denote pixel locations, $d \in \{x, y, z\}$ denotes an axis, and $[G(\vec{p})]_d$ denotes the d -component
 136 of $G(\vec{p})$. Moreover, instead of regressing per-pixel displacements, we predict a matrix Ψ of spatial gradients
 137 between consecutive pixels along each axis. The actual grid G can then be obtained by applying an
 138 integration operation on Ψ along the x -, y - and z -axes, which is approximated by the cumulative sum in
 139 the discrete case. Consequently, two pixels \vec{p} and $\vec{p} + 1$ will have moved closer, maintained distance, or
 140 moved apart in the warped image, if $\Psi_{\vec{p}}$ is respectively less than 1, equal to 1, or greater than 1.

141 2.4 Network Architecture

142 Our network architecture is a modified version of the fully convolutional VNet (Milletari et al., 2016)
 143 for the underlying encoder and decoders parts, maintaining the depth of the model and the rest of the
 144 filter’s configuration unchanged. The model, whose computational graph is displayed in Table 1, comprises
 145 several sequential residual convolutional blocks made of one to three convolutional layers, followed by
 146 downsampling convolutions for the encoder part and upsampling convolutions for the decoder part. We
 147 replaced the initial $5 \times 5 \times 5$ convolutions filter-size by $3 \times 3 \times 3$ in order to reduce the number of parameters
 148 without changing the depth of the model, and also replace PReLU activations by ReLU ones. In order to
 149 speed up its convergence, the model uses residual connections between each encoding and corresponding
 150 decoding stage for both the segmentation and the registration decoder. This allows every layer of the
 151 network, particularly the first ones, to be trained more efficiently since the gradient can flow easier from
 152 the last layers to the first ones with less vanishing or exploding gradient issues. The encoder part deals with
 153 4-inputs per volume, representing the 4 different MRI modalities that are available on the BraTS dataset,
 154 an extra $1 \times 1 \times 1$ convolution is added to fuse the initial modalities. Moreover, the architecture contains 2
 155 decoders of identical blocks, 1 dedicated to the segmentation of tumors for the source and reference image
 156 and 1 dedicated to the optimal displacement that will map the source to the reference image.

157 2.5 Optimization

158 The network is trained to minimize the segmentation and registration loss functions jointly. For the
 159 segmentation task the loss function is summarized in Eq. 1. For registration, the classical optimization
 160 scheme is to minimize the Frobenius norm between the R and D image intensities:

$$\mathcal{L}_{reg} = \|(R - D)\|^2 + \alpha \|\Psi - \Psi_I\|_1 \tag{2}$$

161 Here, in order to better achieve overall registration, the Frobenius norm within the regions predicted to
 162 be tumors is excluded from the loss function. We argue that by doing this, the model does not focus on
 163 tumor regions, which might produce very high norm due to their texture, but rather focuses on the overall
 164 registration task by looking at regions outside the tumor which contain information more pertinent to the
 165 alignment of the volumes. Here we should mention that on \hat{S}_{seg} we apply the same displacement grid as on
 166 S , resulting in $D_{seg} = \mathcal{W}(\hat{S}_{seg}, G)$. Further, let \hat{R}_{seg}^0 and D_{seg}^0 be binary volumes indicating the voxels
 167 which are predicted to be outside any segmented regions. Then, the registration loss can be written as

$$\mathcal{L}_{reg}^* = \|(R - D) \cdot D_{seg}^0 \cdot \hat{R}_{seg}^0\|^2 + \alpha \|\Psi - \Psi_I\|_1 \tag{3}$$

168 where \cdot is the element-wise multiplication, $\|\cdot\|^2$ indicates the Frobenius norm, Ψ_I is the spatial gradient
 169 of the identity deformation and α is the regularization hyperparameter. The use of regularisation on the
 170 displacements Ψ is essential in order to constrain the network to predict smooth deformation grids that are
 171 anatomically more meaningful while at the same time regularize the objective function towards avoiding
 172 local minimum.

173 Finally the final optimisation of the framework is performed by the joint optimisation of the segmentation
 174 and registration loss functions

$$\mathcal{L} = \mathcal{L}_{reg} + \beta \mathcal{L}_{seg}$$

175 where β is a weight that indicates the influence of each of the components on the joint optimization of
 176 the network and was defined after grid search.

177 For the training process, the initial learning rate was $2 \cdot 10^{-3}$ and subdued by a factor of 5 if the
 178 performance on the validation set did not improve for 30 epochs. The training procedure stops when there
 179 is no improvement for 50 epochs. The regularization weights α and β were set to 10^{-10} and 1 after grid
 180 search. As training samples, random pairs among all cases were selected with a batch size limited to 2
 181 due to the limited memory resources on the GPU. The performance of the network was evaluated every
 182 100 batches, and both proposed models converged after nearly 200 epochs. The overall training time was
 183 calculated to ~ 20 hours, while the time for inference of one pair, using 4 different modalities was ~ 3 sec,
 184 using an NVIDIA GeForce GTX 1080 Ti GPU.

185 2.6 Datasets

186 We evaluated the performance of our method using two publicly available datasets, namely the Brain
 187 Tumor Segmentation (BraTS) (Bakas et al., 2018) and Open Access Series of Imaging Studies (OASIS
 188 3) (Marcus et al., 2010) datasets. BraTS contains multi-institutional pre-operative MRI scans of whole
 189 brains with visible gliomas, which are intrinsically heterogeneous in their imaging phenotype (shape and
 190 appearance) and histology. The MRIs are all pre-operative and consist of 4 modalities, i.e. 4 3D volumes,
 191 namely a) a native T1-weighted scan (T1), b) a post-contrast Gadolinium T1-weighted scan (T1Gd), c) a
 192 native T2-weighted scan (T2), and d) a native T2 Fluid Attenuated Inversion Recovery scan (T2-FLAIR).
 193 The BraTS MRIs are provided with voxelwise ground-truth annotations for 5 disjoint classes denoting a)
 194 the background, b) the necrotic and non-enhancing tumor core (NCR/NET), c) the GD-enhancing tumor
 195 (ET), d) the peritumoral edema (ED) as well as invaded tissue, and finally e) the rest of the brain, i.e. brain
 196 with no abnormality nor invaded tissue. Each center was responsible for annotating their MRIs, with a
 197 central validation by domain experts. We use the original dataset split of BraTS 2018 which contains 285
 198 training samples and 66 for validation. In order to perform our experiments, we split this training set into 3
 199 parts, i.e. train, validation and test sets (199, 26 and 60 patients, respectively), while we used the 66 unseen
 200 cases on the platform to report the performance of the proposed and the benchmarked methods. Moreover,
 201 and especially for the registration task, we evaluated the performance of the models trained on BraTS on the
 202 OASIS 3 dataset to test the generalisation of the method. This dataset consists of a longitudinal collection
 203 of 150 subjects which were characterized as either nondemented or with mild cases of Alzheimer’s disease
 204 (AD) using the Clinical Dementia Rating (CDR). Each scan is made of 3 to 4 individual T1-weighted
 205 MRIs, which has been intended to reduce the signal-to-noise ratio visible with single images. The scans
 206 are also provided with annotations for 47 different structures for left and right side of the brain generated
 207 with FreeSurfer. Some samples of both datasets can be seen in Figure 2.

208 The same pre-processing steps have been applied for both datasets. MRIs were resampled to voxels
209 of volume $1mm^3$ using trilinear interpolation. Each scan is then centered by automatically translating
210 their barycenter to the center of the volume. Ground-truth masks of training and validation steps were
211 accordingly translated. Each modality of each scan has been standardised, i.e. the values of the voxels of
212 the 3D subscans were of zero mean and of unit variance. This normalization step is done independently for
213 each patient and for each channel in order to equally consider each channel since modalities have voxels
214 values in completely different ranges. Finally, these consequent scans are cropped into (144, 208, 144)
215 sized volumes.

216 2.7 Statistical evaluations

217 Our contributions in the study are threefold: multi-task segmentation and registration, registration with a
218 shared encoder and latent space merge operator, as well as the loss \mathcal{L}_{reg}^* (Equation 3) that alleviates the
219 registration modifications of tumor tissues in both source and reference patients. Our experiments were
220 intended to weigh the impact of these novelties for both tumor segmentation and registration of MRIs with
221 tumor areas.

222 2.7.1 Methods benchmarked

223 We therefore benchmark multiple versions of our proposed approach with a subset of these novelties to
224 assess their impact on both registration and segmentation. We notably derive 2 variants for both merging
225 operators subtraction and concatenation. The first variant is our fully proposed architecture with a shared
226 encoder for registration and one decoder for segmentation whose tumor predictions are used to implement
227 the proposed loss \mathcal{L}_{reg}^* . These models are named "Proposed concatenation with \mathcal{L}_{reg}^* " and "Proposed
228 subtraction with \mathcal{L}_{reg}^* ". The second variant of models does not use the proposed loss, and are identified
229 with "w/o \mathcal{L}_{reg}^* ". Finally, we also derive a third variant of our approach, yielding one method per merging
230 operator, by discarding the segmentation decoder. Because the proposed loss use the predicted tumor maps
231 from a segmentation decoder, this variant does not rely on it. These latter methods are named "Proposed
232 concatenation only reg." and "Proposed subtraction only reg.", and are primarily benchmarked to assess the
233 performance of the segmentation decoder and the loss \mathcal{L}_{reg}^* with respect to our fully proposed architecture.

234 We also benchmark baseline methods, without any of the proposed contributions. Since our deep learning
235 architecture is derived from the Vnet (Milletari et al., 2016), this model is used as baseline for segmentation.
236 This comparison seems fair since the fully proposed approach can be seen as a Vnet for the task of
237 segmentation: the shared encoder and the proposed loss are primarily designed for registration, and have
238 no direct impact on the segmentation apart from the features learnt in the encoder. For completeness, the
239 top performing results on the BraTS (Bakas et al., 2018) challenge are reported, although we argue that
240 the comparison is unfair since our deep learning architecture is entirely based on the Vnet (Milletari et al.,
241 2016), which is not specifically designed to perform well on the BraTS segmentation task. Finally, we also
242 report the performance of Voxelmorph (Dalca et al., 2018), a well performing brain MRI registration neural
243 network-based approach, although their entire deep learning structure as well as their grid formulation is
244 different.

245 2.7.2 Performance assessment

246 For performance assessment of the segmentation task, we reported the Dice coefficient metric and
247 Hausdorff distance to measure the performance for the tumor classes Tumor Core (TC), Enhancing Tumor
248 (ET) and Whole Tumor (WT) as computed and provided from the BraTS submission website. These classes
249 are the ones used in the BraTS challenge (Bakas et al., 2018), but differ from the original ones provided in

250 the BraTS dataset: TC is the same as the one labelled in the BraTS dataset for necrotic core (NCR/NET),
251 ET is the disjoint union of the original classes NCR/NET and ET, while WT refers to the union of all
252 tumoral and invaded tissues.

253 For the registration, we evaluated the change on the tumor area together with the Dice coefficient
254 metric for the following categories of the OASIS 3 dataset: brain stem (BS), cerebrospinal fluid (CSF),
255 4th ventricle (4V), amygdala (Am), caudate (Ca), cerebellum cortex (CblmC), cerebellum white matter
256 (CblmWM), cerebral cortex (Ceb1C), cerebral white matter (Ceb1WM), hippocampus (Hi), lateral ventricle
257 (LV), pallidum (Pa), putamen (Pu), ventral DC (VDC) and 3rd ventricle (3V) categories. Here we should
258 mention that for the experiments with the OASIS 3 dataset, we performed a training only with the T1-
259 weighted MRIs of the BraTS dataset, in order to match the available modalities of the OASIS 3 dataset.
260 This evaluation is important as *i)* BraTS does not provide anatomical annotations in order to evaluate
261 quantitatively the registration performance and *ii)* the generalisation of the proposed method on an unseen
262 dataset is evaluated. For the registration of tumor tissues, which might not exist in the source or reference
263 MRIs, we expect the model to register tumor areas while maintaining their geometric properties. In
264 particular, we do not really expect the tumor areas to stay completely unchanged. However, we expect
265 that the volume of the different tumor types would change with a ratio similar to the one that the entire
266 source to the reference volume changes. We calculate this ratio by computing $\frac{D_{seg}^j}{S_{seg}^j}$ where $j = \{0, 1, 2, 3\}$
267 corresponds to the entire brain and the different tumor classes (NCR/NET, ET and ED). We then assess
268 the change of the tumor by calculating the absolute value of the difference between $j = 1$ and every other
269 tumor class. Ideally, we expect a model which preserves the tumor geometry and shape during inference to
270 present a zero difference between the entire brain and tumor class ratio. We independently calculate this
271 difference for each tumor class in order to monitor the behavior of each class, but also after merging the
272 entire tumor area.

273 For statistical significance evaluations between any two methods, we compute independent t-tests as
274 presented in (Rouder et al., 2009), defining as null hypothesis the evaluation metrics of the two populations
275 to be equal. We then report the associated p-value, and the Cohen’s d (Rice and Harris, 2005), which we
276 use to measure the effect size. Such statistical significance evaluation is reported in the form $(t(n); p; d)$
277 where n is the number of samples for each population, $t(n)$ is the t-value, p is the p-value and d is Cohen’s
278 d. We defined the difference of two population means is statistically significant if the associated p-value is
279 lower than 0.005, and consider, as a rule of thumb, that a value of d of 0.20 indicates small effect size, 0.50
280 for medium effect size and 0.80 for large effect size. All of the results in this paper have been computed on
281 unseen testing sets, and the performance of all benchmarked models has been assessed once.

282 For rigor and for each t-test conducted, we ensure the following assumptions are met by the underlying
283 distributions: observations are independent and identically distributed, the outcome variable follows a
284 normal distribution in the population (with (Jarque and Bera, 1980)), and the outcome variable has equal
285 standard deviations in two considered (sub)populations (using Levene’s test (Schultz, 1985)). Finally, when
286 comparing two populations, each made of several subpopulations, we merge such subpopulations into a
287 single set, then compute t-tests on the obtained two gathered-populations.

3 RESULTS

3.1 Evaluation of the Segmentation

Segmentation results for the tumor regions are displayed in Table 2 for the case of the same autoencoder architecture trained only with a segmentation decoder (Baseline *segmentation*) and the proposed method using different merging operations and with or without \mathcal{L}_{reg}^* . One can observe that all evaluated methods perform quite similarly with Dice higher than 0.66 for all the classes and models. The *baseline segmentation* model reports slightly better average Dice coefficient and average Hausdorff distance measurements, with an average Dice 0.03 higher, and an average Hausdorff95 distance 0.6 higher than the proposed with concatenation merging operator, although none of these differences are found statistically significant as indicated in Table 5. As an illustration, for Dice, the minimum received p-value was $p = 0.24$, reported between *baseline segmentation* and *proposed concatenation with \mathcal{L}_{reg}^** together with an associated Cohen's $d = 0.21$ indicating a small size effect. Similarly, for Hausdorff95, the minimum received p-value was $p = 0.46$, reported this time between *baseline segmentation* and *proposed concatenation w/o \mathcal{L}_{reg}^** with $d = 0.13$ also indicating a small size effect, which indicated that the means differences between those two models and any other two models are not statistically significant. This is very promising if we take into account that our proposed model is learning a far more complex architecture addressing both registration and segmentation, with the same volume of training data without significant drop of the segmentation performance.

The superiority of the *baseline segmentation* seems to be presented mainly due to higher performance for the TC class (*baseline segmentation* and *proposed subtraction with \mathcal{L}_{reg}^** : $t(66) = 1.41$; $p = 0.16$; $d = 0.24$). Moreover, the concatenation operation seems to perform slightly better for the tumor segmentation than the subtraction, with at least 0.02 improvement for average Dice coefficient, although this improvement is not statistically significant (*proposed concatenation with \mathcal{L}_{reg}^** and *proposed subtraction with \mathcal{L}_{reg}^** : $t(66) = 0.62$; $p = 0.53$; $d = 0.11$).

Moreover, even if one of the main goals of our paper is the proper registration of the tumoral regions, we perform a comparison with the two best performing methods presented in BraTS 2018 (Myronenko, 2018; Isensee et al., 2018) evaluated on the validation dataset of BraTS 2018. In particular, the (Myronenko, 2018) reports an average dice of 0.82, 0.91 and 0.87 for ET, WT and TC respectively, while (Isensee et al., 2018) reports 0.81, 0.91 and 0.87. Both methods outperform our proposed approach on the validation set of BraTS 2018 by integrating novelties specifically designed to the tumor segmentation task of BraTS 2018. In this study, we based our architecture in a relatively simple and widely used 3D fully convolutional network (Milletari et al., 2016) although different architectures with tumor specific components (trained on the evaluated tumor classes), trained on more data (similar to the ones that are used from (Isensee et al., 2018)), or even integrating post processing steps can be easily integrated boosting considerably the performance of our method.

Finally, in Figure 3 we represent the ground truth and predicted tumor segmentation maps comparing the *baseline segmentation* and our proposed method using the different components and merging operators. We present three different cases, two from our custom test set, on which we have the ground truth information and one from the validation set of the BraTS submission page. One can observe that all the methods provide quite accurate segmentation maps for all the three tumor classes.

327 **3.2 Evaluation of the Registration**328 **3.2.1 Evaluation on anatomical structures**

329 The performance of the registration has been evaluated on an unseen dataset with anatomical information,
 330 namely OASIS 3. In Table 3 the mean and standard deviation of the Dice coefficient for the different
 331 evaluated methods are presented. With rigid we indicate the Dice coefficient after the translation of the
 332 volumes such that the center of the brain mass is placed in the center of the volume. It can be observed that
 333 the performance of the evaluated methods are quite similar something which indicates that the additional
 334 tumor segmentation decoder does not decrease the performance of the registration. On the other hand, it
 335 provides additional information about the areas of tumor in the image. From our experiments, we show
 336 that the proposed formulation can provide registration accuracy similar to the recent state-of-the-art deep
 337 learning based methods (Dalca et al., 2018) with approximate the same average Dice values, that is 0.50
 338 for (Dalca et al., 2018) and 0.49 for all but one of the proposed variants. Moreover, again this difference
 339 in the performance between (Dalca et al., 2018) and the proposed method is not statistically significant
 340 with $t(150) = 0.64$; $p = 0.52$; $d = 0.07$. From our comparisons, the only significant difference on the
 341 evaluation of the registration task was reported between the proposed method *concatenation only reg.* with
 342 an average difference of dice reaching 0.05% and with maximum p-values calculated with *concatenation*
 343 *with \mathcal{L}^** ($t(200) = 3, 33$; $p < 10^{-3}$; $d = 0, 38$). From our experiments we saw that the merging operation
 344 affects a lot the performance of the *only reg.* model, with the concatenation reporting the worst average
 345 dice than the rest of the methods.

346 In Figure 4 we present some qualitative evaluation of the registration component, by plotting three
 347 different pairs and their registration from all the evaluated models. The first two columns of the figure
 348 depict the source and reference volumes together with their tissue annotations. The rest of the columns
 349 present the deformed source volume together with the deformed tissue annotations for each of the evaluates
 350 methods. Visually, all methods perform well on the overall shape of the brain with the higher errors in the
 351 deformed annotations being presented at the cerebral white matter and cerebral cortex classes.

352 Finally, we should also mention that the subjects of the OASIS 3 dataset do not contain regions with
 353 tumors. However, our proposed formulation provides tumor masks so that we could evaluate the robustness
 354 of the segmentation part. Indeed, our model for all the different combinations of merging operations
 355 and loss functions, reported a precision score of more than 0.999, indicating its robustness for the tumor
 356 segmentation task.

357 **3.2.2 Evaluation on the tumor areas**

358 Even if the proposed method reports very similar performance with models that perform only registration,
 359 we argue that it addresses better the registration of the tumor areas, maintaining their geometric properties,
 360 as can be inferred in Table 4. This statement is also supported by the statistical tests we performed to
 361 evaluate the difference in performance between the methods, while registering tumor areas (Table 6). In
 362 particular, for each of the tumor classes NCR/NET, ET and ED the difference between the (Dalca et al.,
 363 2018) and the proposed method *subtraction with \mathcal{L}_{reg}^** was significant with NCR/NET: $t(200) = 10.69$;
 364 $p < 10^{-3}$; $d = 1.07$ — ET: $t(200) = 10.51$; $p < 10^{-3}$; $d = 1.05$ — ED: $t(200) = 8.05$; $p < 10^{-3}$;
 365 $d = 0.81$. The similar behavior was obtained when the evaluation was performed by merging all 3 tumor
 366 classes into one (denoted *Combined*). Again, we reported significant differences between (Dalca et al.,
 367 2018) and the proposed method: $t(200) = 11.38$; $p < 10^{-3}$; $d = 1.14$.

368 To evaluate the performance of the different variants of our proposed method, we compared the
369 performance of the proposed *subtraction with \mathcal{L}_{reg}^** and *concatenation with \mathcal{L}_{reg}^** that reported the best
370 performances. Indeed, we did not find significant changes between the two different components except the
371 edema class ($t(200) = 2.78$; $p < 10^{-3}$; $d = 0.28$). Moreover, the proposed *concatenation only reg.* reports
372 also competitive results without using the segmentation masks. In particular, even if the specific method
373 does not report very good performance on the registration evaluated on anatomical structures (Section
374 3.2.1), it reports very competitive performance on the *Combined* and the smallest in size tumor class (*ET*).
375 However, for the other two classes the difference on the performance that it reports in comparison to the
376 proposed variant *subtraction with \mathcal{L}_{reg}^** is significant different: NCR/NET: $t(200) = 6,03$; $p < 10^{-3}$;
377 $d = 0,60$ — ED: $t(200) = 7,03$; $p < 10^{-3}$; $d = 0,70$). Here we should mention that even though
378 *subtraction only reg.* works very well for the registration of the anatomical regions (Section 3.2.1), it
379 reports one of the worst results about tumor preservation, with values close to the ones reported by (Dalca
380 et al., 2018). This indicates again that the *only reg.* model is highly sensitive to the merging operation and
381 it cannot simultaneously provide good performance on tumor areas and registration of the entire volume,
382 proving its inferiority to the proposed method using the *with \mathcal{L}_{reg}^** .

383 Independently of the merging operation with both registration and segmentation tasks, ie with or without
384 \mathcal{L}_{reg}^* , we find that the proposed approach works significantly better in preserving tumor areas when
385 optimized with \mathcal{L}_{reg}^* than without (NCR/NET: $t(200) = -14.33$; $p < 0.005$; $d = 1.43$ — ET: $t(200) =$
386 -9.99 ; $p < 0.005$; $d = 1.00$ — ED: $t(200) = -14.17$; $p < 0.005$; $d = 1.42$ — Combined: $t(200) =$
387 -10.94 ; $p < 0.005$; $d = 1.09$).

388 Figure 5 presents some qualitative examples from the BraTS 2018 to evaluate the performance of the
389 different methods. The first two columns present the pair of images to be registered and segmented and the
390 rest of the columns the deformed source image with the segmented tumor region superimposed. One can
391 observe that the most of the methods that are based only on registration ((Dalca et al., 2018), proposed
392 concatenation and subtraction *only reg.*) together with the proposed concatenation and subtraction *w/o*
393 \mathcal{L}_{reg}^* do not preserve the geometry of the tumor, tending to significantly reduce the area of tumor after
394 registration, or intermix the different types of tumor. On the other hand the behavior of the proposed *with*
395 \mathcal{L}_{reg}^* seems to be much better, with the tumor area properly maintained in the deformed volume.

396 Moreover, in Figure 6 we provide a better visualisation for the displacement grid inside the tumor area,
397 highlighting the importance of Eq. 2. Indeed, one can observe that the displacements inside the tumor area
398 are much smoother and relaxed when we use the information about the tumor segmentation.

4 DISCUSSION

399 In this study, we proposed a novel deep learning based framework to address simultaneously segmentation
400 and registration. The framework combines and generates features, integrating valuable information from
401 both tasks within a bidirectional manner, while it takes advantage of all the available modalities, making
402 it quite robust and generic. The performance of our model indicates highly promising results that are
403 comparable to recent state-of-the-art models that address each of the tasks separately (Dalca et al., 2018).
404 However, we reported a better behavior of the model in the proximity of tumor regions. This behavior has
405 been achieved by training a shared encoder that generates features that are meaningful for both registration
406 and segmentation problems. At the same time, these two problems have been coupled in a joint loss
407 function, enforcing the network to focus on regions that exist in both volumes.

408 Even if we could not do a proper comparison with (Parisot et al., 2012) which shares similar concepts, our
409 method provides very good improvements. In particular, we train both problems at the same time, without
410 using pre-calculated classification probabilities. The method proposed in (Parisot et al., 2012)) is based on
411 a pre-calculated classifier indicating the tumoral regions. The authors provided their segmentation results
412 by adapting Gentle Adaboost algorithm and using different features including intensity values, texture such
413 as Gabor filters and symmetry. After training the classifier they defined an MRF model to optimise their
414 predictions by taking into account pairwise relations. By adopting this strategy, the used probabilities for
415 the tumoral regions are not optimised simultaneously with the registration, something that it is not the case
416 in our methodology. In particular, by sharing representation between the registration and segmentation
417 tasks we argue that we can create features that are more complex and useful sharing information that comes
418 from both problems. By using a deep learning architecture that is end-to-end trainable, we are able to
419 extract features that are suitable to deal with both problems automatically. Moreover, our implementation
420 is modular and scalable permitting easy integration of multiple modalities, something that is not so
421 straightforward with (Parisot et al., 2012) as it is more complicated to adapt and calculate the different
422 similarity measures and classifiers taking into account all these modalities. Finally, we should mention
423 that our method takes advantage of GPU implementation needing only a few seconds in order to provide
424 segmentation and displacement maps while the method in (Parisot et al., 2012) needs approximately 6
425 minutes.

426 Both qualitative and quantitative evaluations of the proposed architecture highlight the great potentials
427 of the proposed method reporting more than 0.66 Dice coefficient for the segmentation of the different
428 tumor areas, evaluated on the publicly available BraTS 2018 validation set. Our formulation reported
429 similar behavior than the model with only the segmentation block which indicates that the joint formulation
430 did not really affect the performance of the tumor segmentation, however, it provides more complex
431 models providing tumor segmentation masks for two images at the same time, predicting simultaneously
432 optimal displacements between them. Moreover, both concatenation and subtraction operators report
433 similar performances, an expected result for the specific segmentation task, since the merging operation is
434 mainly used on the registration decoder, even if it affects the learned parameters of the encoder and thus
435 indirectly the segmentation decoder.

436 Concerning the comparison between top performing tumor segmentation methods, although our
437 formulation underperforms the winning methods of BraTS 2018, we want to highlight two major points.
438 First of all, our formulation is modular in the sense that different network architectures with optimised
439 components for tumor segmentation can be evaluated depending on the application and the goals of the
440 problem. For our experiments we chose a simple VNet architecture (Milletari et al., 2016) proving that
441 the registration components do not significantly hinder the segmentation performance and indicating the
442 soundness of our method however any other encoder decoder architecture can be used and evaluated.
443 Secondly, the main goal of our method was the proper registration and segmentation of the tumoral regions
444 together with the rest of the anatomical structures and that was the main reason we did not optimize
445 our network architecture according to the winning methods of BraTS 2018. However, we demonstrated
446 that with a very simple architecture, we can register properly tumoral and anatomical structures while
447 segmenting with more than 76% of Dice the tumoral regions.

448 Continuing with the evaluation of the registration performance, once more the joint multi task framework
449 reports similar and without statistical difference performance with formulations that address only the
450 registration task evaluated on anatomical regions that exist on both volumes. However, we argue that
451 abnormal regions registration is better addressed both in terms of qualitative and quantitative metrics.

452 Moreover, from our experiments we observed that subtraction of the coding features of the tumors reports
453 higher performances for the registration of the tumor areas. This indicates that the subtraction can capture
454 and code more informative features for the registration task. What is more, we achieved very good
455 generalization for all the deep learning based registration methods, as they reported very stable performance
456 in a completely unseen dataset (part of the OASIS3).

457 Even if, from our experiments, the competence of our proposed method for both registration and
458 segmentation tasks is indicated, we report a much better performance for the registration of the tumoral
459 regions. In particular, in one joint framework we were able to produce efficiently and accurately tumor
460 segmentation maps for both source and reference images together with their displacement maps that register
461 the source volume to the reference volume space. Our experiments indicated that the proposed method
462 with the \mathcal{L}_{reg}^* variant register properly the anatomical together with the tumoral regions with statistical
463 significance compare to the rest of the methods for the latter. Both qualitative and quantitative evaluations
464 of the different components indicate the superiority of the with \mathcal{L}_{reg}^* variant of the proposed method for
465 brain MRI registration with tumor extent preservation. Using such a formulation, the network focus on
466 improving local displacements on tissues anywhere in the common brain space instead of minimizing
467 the loss within the tumoral regions, which are empirically the regions with the highest registration errors.
468 Consequently, the network improves its registration performance on non-tumor regions (as discussed in
469 Section Evaluation on anatomical structures), while also relaxing the obtained displacements inside those
470 predicted tumor regions.

471 Some limitations of our method include the number of parameters that have to be tuned during the
472 training due to the multi task nature of our formulation, namely α and β that affect the performance of the
473 network. Moreover, due to the multimodal nature of the input and the two decoders, the network cannot be
474 very deep due to GPU memory limitations.

475 Although the pipeline was built using different patients for the registration task as a proof of concept,
476 such tool could have numerous applications in clinical practice, especially when applied in different images
477 acquired from the same patient. Regarding the radiotherapy treatment planning, several studies have
478 shown that significant changes of the targeted volumes in the brain occurred during radiotherapy raising
479 the question of replanning treatment to reduce the amount of healthy brain irradiated in case of tumor
480 reduction, or to re-adapt the treatment for brain tumors that grow during radiation (Champ et al., 2012;
481 Yang et al., 2016; Mehta et al., 2018). Since MR-guided linear accelerator will offer the opportunity to
482 acquire daily images during RT treatment, the proposed tool could help with automatic segmentation and
483 image registration for replanning purposes, and it could also allow accurate evaluation of the dose delivered
484 in targeted volumes and healthy tissues by taking into account the different volume changes. Moreover,
485 while changes of imaging features under treatment is known to be associated with treatment outcomes
486 in several cancer diseases (Fave et al., 2017; Vera et al., 2014), the registration grid computed from two
487 same-patient acquisitions realized at different times allows an objective and precise evaluation of the tumor
488 changes.

489 Future work involves a better modeling of the prior knowledge through a more appropriate geometric
490 modeling of tumor proximity that encodes more accurately the registration errors in these areas. This
491 modeling can be integrated into the existing formulation with some additions specific to tumor losses that
492 will further constrain its change. Moreover, we have noticed that the use of Fobenius norm during the
493 training of the registration part is very sensitive to artifacts in the volume, preventing the network process
494 from being completely robust. In the future, we aim to evaluate the performance of the proposed framework

495 using adversarial losses in order to better address multimodal cases. Finally, means to automatically obtain
496 the training parameters α and β would be investigated.

CONFLICT OF INTEREST STATEMENT

497 The authors declare that the research was conducted in the absence of any commercial or financial
498 relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

499 TE, ML, MV, NP, and ED: designed research; TE, ML, and MV: performed research; TE, ML, and MV:
500 analyzed and interpreted data; TE, ML, and MV: wrote the paper; TE, ML, MV, EA, EB, AC, SCha, SChr,
501 MS, RS, CR, HT, NP and ED: revised and approved the paper.

FUNDING

502 This work have been partially funding by the ARC: Grant SIGNIT201801286, the Fondation pour la
503 Recherche Médicale: Grant DIC20161236437, SIRIC-SOCRATE 2.0, ITMO Cancer, Institut National du
504 Cancer (INCa) and Amazon Web Services (AWS).

ACKNOWLEDGMENTS

505 We would like to acknowledge Y. Boursin, M. Azoulay and GustaveRoussy Cancer Campus DTNSI team
506 for providing the infrastructure resources used in this work as well as Amazon Web Services for their
507 partial support.

REFERENCES

- 508 Avants, B., Epstein, C., Grossman, M., and Gee, J. (2008). Symmetric diffeomorphic image registration
509 with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical*
510 *Image Analysis* 12, 26 – 41. doi:<https://doi.org/10.1016/j.media.2007.06.004>. Special Issue on The
511 Third International Workshop on Biomedical Image Registration – WBIR 2006
- 512 Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., et al. (2017a). Advancing the cancer
513 genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific*
514 *data* 4
- 515 Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., et al. (2017b). Segmentation labels
516 and radiomic features for the pre-operative scans of the tcga-gbm collection. *The Cancer Imaging*
517 *Archive*
- 518 Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., et al. (2017c). Segmentation labels
519 and radiomic features for the pre-operative scans of the tcga-lgg collection. *The Cancer Imaging Archive*
- 520 Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., et al. (2018). Identifying the best
521 machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival
522 prediction in the brats challenge. *arXiv preprint arXiv:1811.02629*
- 523 C. Holland, E. (2002). Progenitor cells and glioma formation 14, 683–8
- 524 Champ, C. E., Siglin, J., Mishra, M. V., Shen, X., Werner-Wasik, M., Andrews, D. W., et al. (2012).
525 Evaluating changes in radiation treatment volumes from post-operative to same-day planning mri in
526 high-grade gliomas. *Radiation Oncology* 7, 220

- 527 Chandra, S., Vakalopoulou, M., Fidon, L., Battistella, E., Estienne, T., Sun, R., et al. (2019). Context
528 aware 3d cnns for brain tumor segmentation. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and*
529 *Traumatic Brain Injuries*, eds. A. Crimi, S. Bakas, H. Kuijf, F. Keyvan, M. Reyes, and T. van Walsum
530 (Cham: Springer International Publishing), 299–310
- 531 Christodoulidis, S., Sahasrabudhe, M., Vakalopoulou, M., Chassagnon, G., Revel, M.-P., Mougiakakou, S.,
532 et al. (2018). Linear and deformable image registration with 3d convolutional neural networks. In *Image*
533 *Analysis for Moving Organ, Breast, and Thoracic Images* (Springer). 13–22
- 534 Dalca, A. V., Balakrishnan, G., Guttag, J. V., and Sabuncu, M. R. (2018). Unsupervised learning for fast
535 probabilistic diffeomorphic registration. In *MICCAI*
- 536 Fave, X., Zhang, L., Yang, J., Mackin, D., Balter, P., Gomez, D., et al. (2017). Delta-radiomics features for
537 the prediction of patient outcomes in non-small cell lung cancer. *Scientific reports* 7, 588
- 538 Glocker, B., Komodakis, N., Navab, N., Tziritas, G., and Paragios, N. (2009). Dense registration with
539 deformation priors. In *Information Processing in Medical Imaging*, eds. J. L. Prince, D. L. Pham, and
540 K. J. Myers
- 541 Gooya, A., Biros, G., and Davatzikos, C. (2010). Deformable registration of glioma images using
542 em algorithm and diffusion reaction modeling. *IEEE transactions on medical imaging* 30, 375–90.
543 doi:10.1109/TMI.2010.2078833
- 544 Gooya, A., Pohl, K. M., Bilello, M., Cirillo, L., Biros, G., Melhem, E. R., et al. (2012). Glistr: Glioma
545 image segmentation and registration. *IEEE Transactions on Medical Imaging* 31, 1941–1954
- 546 Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., and Maier-Hein, K. H. (2017). Brain tumor
547 segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge. In *International*
548 *MICCAI Brainlesion Workshop* (Springer), 287–297
- 549 Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., and Maier-Hein, K. H. (2018). No new-net. In
550 *International MICCAI Brainlesion Workshop* (Springer), 234–244
- 551 Jarque, C. M. and Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial
552 independence of regression residuals. *Economics letters* 6, 255–259
- 553 Kamnitsas, K., Bai, W., Ferrante, E., McDonagh, S., Sinclair, M., Pawlowski, N., et al. (2018). Ensembles
554 of multiple models and architectures for robust brain tumour segmentation. In *Brainlesion: Glioma,*
555 *Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, eds. A. Crimi, S. Bakas, H. Kuijf, B. Menze,
556 and M. Reyes (Cham: Springer International Publishing), 450–462
- 557 Klein, S., Staring, M., Murphy, K., A Viergever, M., and P W Pluim, J. (2009). Elastix: A toolbox
558 for intensity-based medical image registration. *IEEE transactions on medical imaging* 29, 196–205.
559 doi:10.1109/TMI.2009.2035616
- 560 Lin, T.-Y., Goyal, P., Girshick, R. B., He, K., and Dollár, P. (2017). Focal loss for dense object detection.
561 *2017 IEEE International Conference on Computer Vision (ICCV)*, 2999–3007
- 562 Marcus, D. S., Fotenos, A. F., Csernansky, J. G., Morris, J. C., and Buckner, R. L. (2010). Open access
563 series of imaging studies: longitudinal mri data in nondemented and demented older adults. *Journal of*
564 *cognitive neuroscience* 22, 2677–2684
- 565 McKinley, R., Meier, R., and Wiest, R. (2018). Ensembles of densely-connected cnns with label-uncertainty
566 for brain tumor segmentation. In *International MICCAI Brainlesion Workshop* (Springer), 456–465
- 567 Mehta, S., Gajjar, S. R., Padgett, K. R., Asher, D., Stoyanova, R., Ford, J. C., et al. (2018). Daily tracking
568 of glioblastoma resection cavity, cerebral edema, and tumor volume with mri-guided radiation therapy.
569 *Cureus* 10

- 570 Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., et al. (2015). The
571 multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging*
572 34
- 573 Milletari, F., Navab, N., and Ahmadi, S.-A. (2016). V-net: Fully convolutional neural networks for
574 volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*
575 (IEEE), 565–571
- 576 Myronenko, A. (2018). 3d MRI brain tumor segmentation using autoencoder regularization. In
577 *International MICCAI Brainlesion Workshop* (Springer), 311–320
- 578 Ou, Y., Sotiras, A., Paragios, N., and Davatzikos, C. (2011). Dramms: Deformable registration via
579 attribute matching and mutual-saliency weighting. *Medical Image Analysis* 15, 622 – 639. doi:<https://doi.org/10.1016/j.media.2010.07.002>. Special section on IPMI 2009
- 581 Parisot, S., Darlix, A., Baumann, C., Zouaoui, S., Yordanova, Y., Blonski, M., et al. (2016). A probabilistic
582 atlas of diffuse who grade ii glioma locations in the brain. *PloS one* 11, e0144200. doi:10.1371/journal.
583 pone.0144200
- 584 Parisot, S., Duffau, H., Chemouny, S., and Paragios, N. (2012). Joint tumor segmentation and dense
585 deformable registration of brain mr images. In *Medical Image Computing and Computer-Assisted*
586 *Intervention – MICCAI 2012*, eds. N. Ayache, H. Delingette, P. Golland, and K. Mori
- 587 Postelnicu, G., Zollei, L., and Fischl, B. (2009). Combined volumetric and surface registration. *IEEE*
588 *Transactions on Medical Imaging* 28
- 589 Rice, M. E. and Harris, G. T. (2005). Comparing effect sizes in follow-up studies: Roc area, cohen’s d, and
590 r. *Law and human behavior* 29, 615–620
- 591 Robinson, E. C., Garcia, K., Glasser, M. F., Chen, Z., Coalson, T. S., Makropoulos, A., et al. (2018).
592 Multimodal surface matching with higher-order smoothness constraints. *NeuroImage* 167
- 593 Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., and Iverson, G. (2009). Bayesian t tests for
594 accepting and rejecting the null hypothesis. *Psychonomic bulletin & review* 16, 225–237
- 595 Schultz, B. B. (1985). Levene’s test for relative variation. *Systematic Zoology* 34, 449–456
- 596 Sepúlveda-Sánchez, J., Langa, J. M., Arráez, M., Fuster, J., Laín, A. H., Reynés, G., et al. (2018). Seom
597 clinical guideline of diagnosis and management of low-grade glioma (2017). *Clinical and Translational*
598 *Oncology* 20, 3–15
- 599 Shi, W., Jantsch, M., Aljabar, P., Pizarro, L., Bai, W., Wang, H., et al. (2013). Temporal sparse free-form
600 deformations. *Medical Image Analysis* 17, 779 – 789. doi:<https://doi.org/10.1016/j.media.2013.04.010>.
601 Special Issue on the 2012 Conference on Medical Image Computing and Computer Assisted Intervention
- 602 Shu, Z., Sahasrabudhe, M., Riza, A. G., Samaras, D., Paragios, N., and Kokkinos, I. (2018). Deforming
603 autoencoders: Unsupervised disentangling of shape and appearance. In *The European Conference on*
604 *Computer Vision (ECCV)*
- 605 Sotiras, A., Davatzikos, C., and Paragios, N. (2013). Deformable medical image registration: A survey.
606 *IEEE Transactions on Medical Imaging* 32, 1153–1190
- 607 Stupp, R., Brada, M., Van Den Bent, M., Tonn, J.-C., and Pentheroudakis, G. (2014). High-grade
608 glioma: Esmo clinical practice guidelines for diagnosis, treatment and follow-up. *Annals of oncology* 25,
609 iii93–iii101
- 610 Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., and Jorge Cardoso, M. (2017). Generalised dice
611 overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in*
612 *Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, eds. M. J. Cardoso,
613 T. Arbel, G. Carneiro, T. Syeda-Mahmood, J. M. R. Tavares, M. Moradi, A. Bradley, H. Greenspan,

- 614 J. P. Papa, A. Madabhushi, J. C. Nascimento, J. S. Cardoso, V. Belagiannis, and Z. Lu (Cham: Springer
 615 International Publishing), 240–248
- 616 Vera, P., Dubray, B., Palie, O., Buvat, I., Hapdey, S., Modzelewski, R., et al. (2014). Monitoring tumour
 617 response during chemo-radiotherapy: a parametric method using fdg-pet/ct images in patients with
 618 oesophageal cancer. *EJNMMI research* 4, 12
- 619 Wang, G., Li, W., Ourselin, S., and Vercauteren, T. (2017). Automatic brain tumor segmentation using
 620 cascaded anisotropic convolutional neural networks. In *International MICCAI Brainlesion Workshop*
 621 (Springer), 178–190
- 622 Wee, C. W., Sung, W., Kang, H.-C., Cho, K. H., Han, T. J., Jeong, B.-K., et al. (2015). Evaluation of
 623 variability in target volume delineation for newly diagnosed glioblastoma: a multi-institutional study from
 624 the korean radiation oncology group. *Radiation Oncology* 10, 137. doi:10.1186/s13014-015-0439-z
- 625 Yang, Z., Zhang, Z., Wang, X., Hu, Y., Lyu, Z., Huo, L., et al. (2016). Intensity-modulated radiotherapy for
 626 gliomas: dosimetric effects of changes in gross tumor volume on organs at risk and healthy brain tissue.
 627 *OncoTargets and therapy* 9, 3545
- 628 Zhao, Z., Yang, G., Lin, Y., Pang, H., and Wang, M. (2018). Automated glioma detection and segmentation
 629 using graphical models. *PLOS ONE* 13, e0200745. doi:10.1371/journal.pone.0200745
- 630 Zhou, C., Chen, S., Ding, C., and Tao, D. (2018). Learning contextual and attentive information for brain
 631 tumor segmentation. In *International MICCAI Brainlesion Workshop* (Springer), 497–507

TABLES

FIGURES

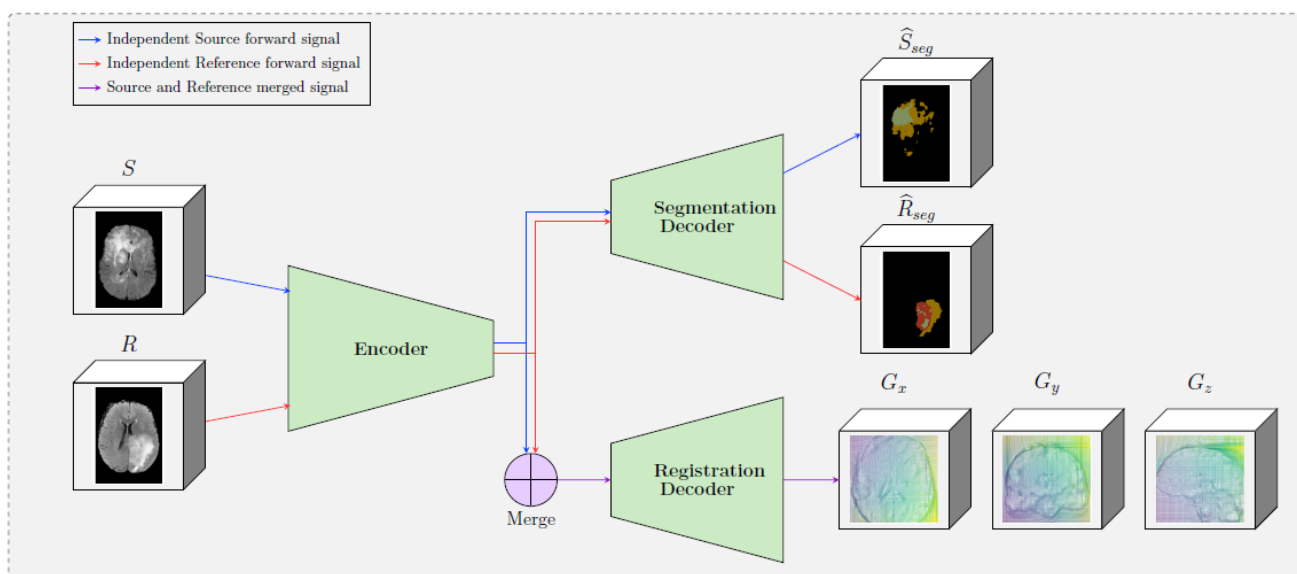


Figure 1. A schematic representation of the proposed framework. The framework is composed by two decoders, one which provides tumor segmentation masks for both S and R images, and one the provides the optimal displacement grid G that will accurately map the S to the R image. The merge bloc will combine the forward signal of the source input and the reference input (which are forwarded independently in the encoder).

Name	Input	Res. input	Operations	Output shape
Encoder				
Enc ¹	4D MRI		Conv _{1,8} , ReLU, (Conv _{3,8} , ReLU), AddId,	(144, 208, 144, 8)
Enc ²	Enc ¹		Conv _{2,16} , ReLU, (Conv _{3,16} , ReLU)*2, AddId	(72, 104, 72, 16)
Enc ³	Enc ²		Conv _{2,32} , ReLU, (Conv _{3,32} , ReLU)*3, AddId	(36, 52, 36, 32)
Enc ⁴	Enc ³		Conv _{2,64} , ReLU, (Conv _{3,64} , ReLU)*3, AddId	(18, 26, 18, 64)
Enc ⁵	Enc ⁴		Conv _{2,128} , ReLU, (Conv _{3,128} , ReLU)*3, AddId	(9, 13, 9, 128)
Segmentation decoder				
Dec ⁴ _{seg}	Enc ⁵	Enc ⁴	DeConv _{2,64} , ReLU, ResConc, (Conv _{3,64} , ReLU)*3, AddId	(18, 26, 18, 64)
Dec ³ _{seg}	Dec ⁴ _{seg}	Enc ³	DeConv _{2,32} , ReLU, ResConc, (Conv _{3,32} , ReLU)*3, AddId	(36, 52, 36, 32)
Dec ² _{seg}	Dec ³ _{seg}	Enc ²	DeConv _{2,16} , ReLU, ResConc, (Conv _{3,16} , ReLU)*2, AddId	(72, 104, 72, 16)
Dec ¹ _{seg}	Dec ² _{seg}	Enc ¹	DeConv _{2,8} , ReLU, ResConc, (Conv _{3,8} , ReLU), AddId	(144, 208, 144, 8)
Dec ⁰ _{seg}	Dec ¹ _{seg}		Conv _{1,4} , Softmax	(144, 208, 144, 4)
Registration decoder				
Merge	Enc ⁱ _R , Enc ⁱ _S		For all $1 \leq i \leq 5$, MEnc ⁱ = Enc ⁱ _R \oplus Enc ⁱ _S	
Dec ⁴ _{reg}	MEnc ⁵	MEnc ⁴	DeConv _{2,64} , ReLU, ResConc, (Conv _{3,64} , ReLU)*3, AddId	(18, 26, 18, 64)
Dec ³ _{reg}	Dec ⁴ _{reg}	MEnc ³	DeConv _{2,32} , ReLU, ResConc, (Conv _{3,32} , ReLU)*3, AddId	(36, 52, 36, 32)
Dec ² _{reg}	Dec ³ _{reg}	MEnc ²	DeConv _{2,16} , ReLU, ResConc, (Conv _{3,16} , ReLU)*2, AddId	(72, 104, 72, 16)
Dec ¹ _{reg}	Dec ² _{reg}	MEnc ¹	DeConv _{2,8} , ReLU, ResConc, (Conv _{3,8} , ReLU), AddId	(144, 208, 144, 8)
Dec ⁰ _{reg}	Dec ¹ _{reg}		Conv _{1,3} , Sigmoid	(144, 208, 144, 3)

Table 1. Layer architecture of the encoder, the segmentation and the registration decoders. The sub-architectures are grouped into blocks, one per table line, whose names are indicated in the first column. Each block processed a forward signal as input identified by the second column. Additionally, both decoders have residual connections from different stages of the encoder, identified by the third column. The blocks are made of a set of successive operations where Conv_{w,f} (resp. DeConv_{w,f}) stands for a convolutional (resp. deconvolutional) layer with weight size $w \times w \times w$ and f filters, ReLU - Rectified Linear Unit, AddId - intra-block residual connection with the output of the first activated convolution of the corresponding block, ResConc - encoder to decoder residual connection from the output of the third column block to the current signal, Softmax and Sigmoid - finale output activation. * indicates successive repetition of the previous operations in parenthesis. For convolutions and deconvolutions layers, strides is $1 \times 1 \times 1$ except for the Conv₂, which is $2 \times 2 \times 2$. The first layer of the registration decoder indicates the merging operation of the source signal and the reference signal, which are obtained by inferring them successively in the encoder; \oplus indicates elementwise subtraction or channelwise concatenation of the source and reference list of tensors (forward network signal and 4 residual connection signals). The last column indicates each block output shape (channels last).

Method	Average		Dice			Hausdorff95		
	Dice	Hausdorff95	ET	WT	TC	ET	WT	TC
Baseline segmentation	0.79 \pm 0.29	7.0 \pm 9.6	0.73 \pm 0.29	0.87 \pm 0.13	0.75 \pm 0.24	4.7 \pm 8.2	7.2 \pm 9.4	9.2 \pm 8.9
Proposed								
concatenation w/o \mathcal{L}_{reg}^*	0.74 \pm 0.29	8.3 \pm 10.4	0.70 \pm 0.29	0.87 \pm 0.11	0.65 \pm 0.29	6.2 \pm 9.8	7.8 \pm 11.1	11.3 \pm 7.1
concatenation with \mathcal{L}_{reg}^*	0.73 \pm 0.29	7.6 \pm 9.9	0.68 \pm 0.30	0.87 \pm 0.12	0.66 \pm 0.28	6.3 \pm 9.9	5.6 \pm 4.2	10.8 \pm 6.6
subtraction w/o \mathcal{L}_{reg}^*	0.76 \pm 0.27	7.8 \pm 10.3	0.71 \pm 0.28	0.88 \pm 0.10	0.70 \pm 0.24	6.5 \pm 10.8	7.4 \pm 11.0	10.0 \pm 7.4
subtraction with \mathcal{L}_{reg}^*	0.76 \pm 0.27	7.9 \pm 10.1	0.71 \pm 0.29	0.88 \pm 0.10	0.69 \pm 0.25	5.8 \pm 9.6	7.7 \pm 11.5	11.1 \pm 8.3

Table 2. Quantitative results of the different methods on the segmentation task on the BraTS 2018 validation dataset. Dice and Hausdorff95 are reported for the three classes Whole Tumor (WT), Enhancing Tumor (ET) and Tumor Core (TC) together with their average values. Results are reported with mean across patients (MRIs) along with the associated standard deviation. We upload our predictions on the official leaderboard of the validation set (66 patients).

Method	BS	CSF	CblmC	CblmWM	CeblWM	Pu	VDC	Pa	Ca	LV	Hi	3V	4V	Am	CeblC	Average
Rigid	0.58 ± 0.15	0.39 ± 0.11	0.46 ± 0.13	0.40 ± 0.14	0.49 ± 0.05	0.44 ± 0.13	0.47 ± 0.13	0.35 ± 0.17	0.27 ± 0.15	0.40 ± 0.13	0.34 ± 0.13	0.39 ± 0.17	0.15 ± 0.15	0.24 ± 0.18	0.36 ± 0.04	0.38 ± 0.13
Voxelmorph	0.69 ± 0.12	0.46 ± 0.13	0.63 ± 0.11	0.57 ± 0.13	0.73 ± 0.083	0.42 ± 0.14	0.5 ± 0.11	0.33 ± 0.14	0.42 ± 0.17	0.62 ± 0.14	0.38 ± 0.13	0.53 ± 0.18	0.32 ± 0.23	0.25 ± 0.17	0.6 ± 0.084	0.5 ± 0.14
Proposed concatenation only reg.	0.65 ± 0.15	0.34 ± 0.1	0.58 ± 0.11	0.48 ± 0.14	0.6 ± 0.056	0.46 ± 0.12	0.47 ± 0.12	0.38 ± 0.14	0.35 ± 0.15	0.54 ± 0.14	0.35 ± 0.13	0.4 ± 0.16	0.21 ± 0.17	0.27 ± 0.18	0.46 ± 0.051	0.44 ± 0.13
w/o \mathcal{L}_{reg}^*	0.72 ± 0.13	0.42 ± 0.1	0.61 ± 0.11	0.51 ± 0.12	0.63 ± 0.056	0.47 ± 0.14	0.51 ± 0.12	0.37 ± 0.16	0.44 ± 0.15	0.65 ± 0.13	0.42 ± 0.14	0.46 ± 0.17	0.31 ± 0.22	0.31 ± 0.19	0.48 ± 0.052	0.49 ± 0.13
with \mathcal{L}_{reg}^*	0.7 ± 0.15	0.44 ± 0.12	0.6 ± 0.13	0.52 ± 0.14	0.66 ± 0.06	0.47 ± 0.14	0.52 ± 0.13	0.38 ± 0.16	0.42 ± 0.16	0.65 ± 0.14	0.4 ± 0.15	0.51 ± 0.19	0.3 ± 0.22	0.28 ± 0.2	0.49 ± 0.058	0.49 ± 0.14
subtraction only reg.	0.71 ± 0.13	0.41 ± 0.1	0.61 ± 0.12	0.53 ± 0.13	0.66 ± 0.058	0.47 ± 0.12	0.5 ± 0.11	0.37 ± 0.15	0.43 ± 0.14	0.63 ± 0.12	0.4 ± 0.13	0.47 ± 0.16	0.34 ± 0.22	0.29 ± 0.19	0.49 ± 0.054	0.49 ± 0.13
w/o \mathcal{L}_{reg}^*	0.7 ± 0.13	0.41 ± 0.1	0.6 ± 0.11	0.52 ± 0.12	0.65 ± 0.057	0.48 ± 0.13	0.53 ± 0.11	0.39 ± 0.15	0.43 ± 0.14	0.64 ± 0.13	0.41 ± 0.13	0.49 ± 0.17	0.3 ± 0.22	0.29 ± 0.18	0.48 ± 0.053	0.49 ± 0.13
with \mathcal{L}_{reg}^*	0.72 ± 0.12	0.4 ± 0.11	0.61 ± 0.11	0.53 ± 0.12	0.64 ± 0.058	0.47 ± 0.12	0.51 ± 0.11	0.38 ± 0.15	0.41 ± 0.15	0.63 ± 0.13	0.43 ± 0.13	0.44 ± 0.17	0.3 ± 0.22	0.33 ± 0.18	0.48 ± 0.054	0.49 ± 0.13

Table 3. The mean and standard deviation of the dice coefficient for the 15 different classes of OASIS 3 dataset for the different evaluated methods. The first two rows are baseline methods. The rest of the rows present the results of our proposed method evaluating the different variants and merging operators. The names of the columns represent various brain structures, namely: brain stem (BS), cerebrospinal fluid (CSF), 4th ventricle (4V), amygdala (Am), caudate (Ca), cerebellum cortex (CblmC), cerebellum white matter (CblmWM), cerebral cortex (CeblC), cerebral white matter (CeblWM), hippocampus (Hi), lateral ventricle (LV), pallidum (Pa), putamen (Pu), ventral DC (VDC) and 3rd ventricle (3V).

Method	NCR/NET	ET	ED	Combined
(Dalca et al., 2018)	2.27 ± 2.68	0.67 ± 0.55	1.96 ± 3.03	0.62 ± 0.51
Proposed concatenation only reg.	0.51 ± 0.61	0.26 ± 0.19	0.71 ± 0.94	0.22 ± 0.15
concatenation w/o \mathcal{L}_{reg}^*	1.35 ± 1.14	0.64 ± 0.41	1.80 ± 1.82	0.64 ± 0.42
concatenation with \mathcal{L}_{reg}^*	0.26 ± 0.20	0.26 ± 0.13	0.30 ± 0.28	0.21 ± 0.12
subtraction only reg.	1.34 ± 0.77	0.77 ± 0.59	2.02 ± 1.65	0.68 ± 0.52
subtraction w/o \mathcal{L}_{reg}^*	1.74 ± 1.35	0.72 ± 0.72	2.38 ± 1.74	0.74 ± 0.76
subtraction with \mathcal{L}_{reg}^*	0.24 ± 0.17	0.25 ± 0.13	0.23 ± 0.22	0.20 ± 0.11

Table 4. Quantitative estimates on tumor shrinking. The measure used is the average over 200 testing pairs of patients of the distance between the ratio of the volumes of the deformed source ground-truth mask and the original ground-truth mask for each original class of the BraTS 2018 dataset (NCR/NET, ET and ED), and the ratio of the reference brain volume over the source brain volume. In this context, the best performance reachable is 0 for each class. Additionally, ground-truth masks are binarized into Whole Tumor masks, with a value of 1 if and only if a voxel is annotated as one of the 3 tumor classes, and the same measure is computed in the last column ("Combined"), which should indicate the overall impact of tumor shrinking of the whole tumor without considering swapping of intra-tumoral classes.

Method	Average		Dice			Hausdorff95		
	Dice	Hausdorff95	ET	WT	TC	ET	WT	TC
Baseline segmentation	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Proposed								
concatenation w/o \mathcal{L}_{reg}^*	0.32	0.46	0.55	1.00	0.03	0.34	0.74	0.14
concatenation with \mathcal{L}_{reg}^*	0.24	0.72	0.33	1.00	0.05	0.31	0.21	0.24
subtraction w/o \mathcal{L}_{reg}^*	0.55	0.65	0.69	0.62	0.24	0.28	0.91	0.58
subtraction with \mathcal{L}_{reg}^*	0.55	0.60	0.69	0.62	0.16	0.48	0.79	0.21

Table 5. Statistical significance of the proposed methods with (Milletari et al., 2016) on the BraTS segmentation task. For each model (line) and each performance measure (column), the displayed value is the p-value, up to 2 significant figures, of the statistical significance between the model and (Milletari et al., 2016) for the corresponding measure (Dice or Hausdorff95) on the corresponding tumor class (ET, WT, TC, or the union of the 3 latter in the two columns *Average*) on the 66 testing samples of BraTS. No p-values are statistically significant between all of the proposed variants and (Milletari et al., 2016). Blue line represents the reference model, red cells indicate no statistical significant p-values (cutoff 0.005) while green color represent statistical significant p-values.

Method	NCR/NET	ET	ED	Combined
(Dalca et al., 2018)	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
Proposed				
concatenation only reg.	$< 10^{-3}$	0.540	$< 10^{-3}$	0.130
concatenation w/o \mathcal{L}_{reg}^*	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
concatenation with \mathcal{L}_{reg}^*	0.282	0.442	0.006	0.386
subtraction only reg.	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
subtraction w/o \mathcal{L}_{reg}^*	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
subtraction with \mathcal{L}_{reg}^*	1.000	1.000	1.000	1.000

Table 6. Statistical significance of the proposed methods and (Dalca et al., 2018), with the best proposed variant *subtraction with \mathcal{L}_{reg}^** regarding the tumor shrinking preservation on the OASIS 3 registration task. For each model (line) and each performance measure (column), the displayed value is the p-value, up to 3 significant figures, of the statistical significance between the model and *subtraction with \mathcal{L}_{reg}^** for the tumor preservation measure on the corresponding tumor class (NCR/NET, ET, ED, and the union of the 3 latter in the column *Combined*) on the 200 testing pairs of OASIS 3. Blue line represents the reference model, red cells indicate no statistical significant p-values while green color represent statistical significant p-values.

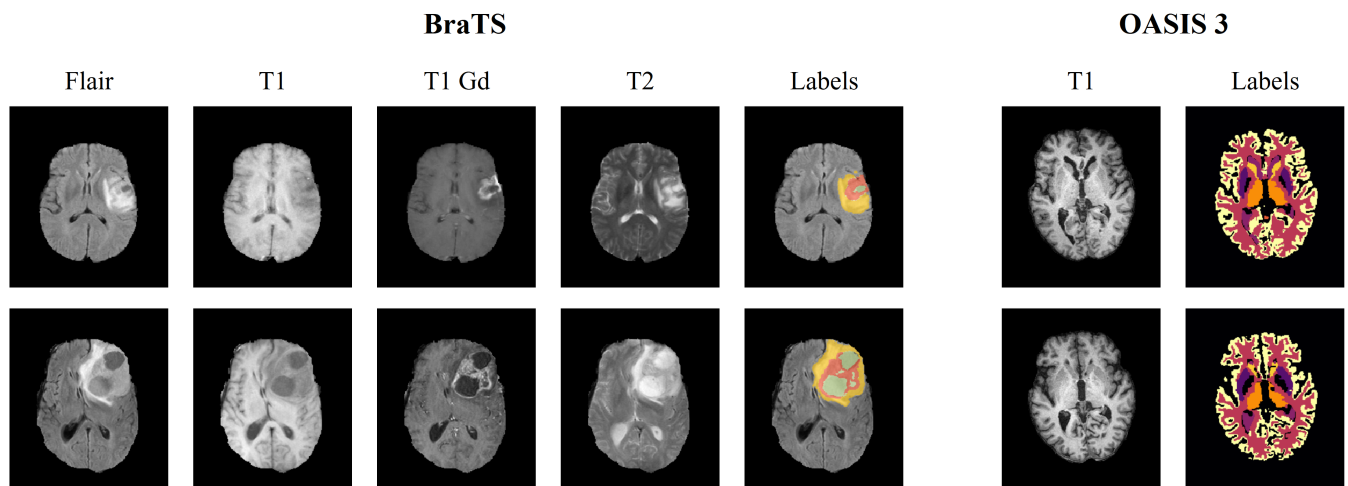


Figure 2. Illustration of one slice from two examples from both BraTS and OASIS 3 datasets. The data from BraTS are 3D spatial volumes with 4 modalities (T1, T1 gadolinium, T2, T2 FLAIR), along with voxelwise annotations for the 3 tumor tissue subclasses depicting the overall extent of tumors. OASIS 3 contains 3D volume only for the T1 modality, and images are provided with voxelwise annotations of 13 normal brain structures for patients without brain tumors.

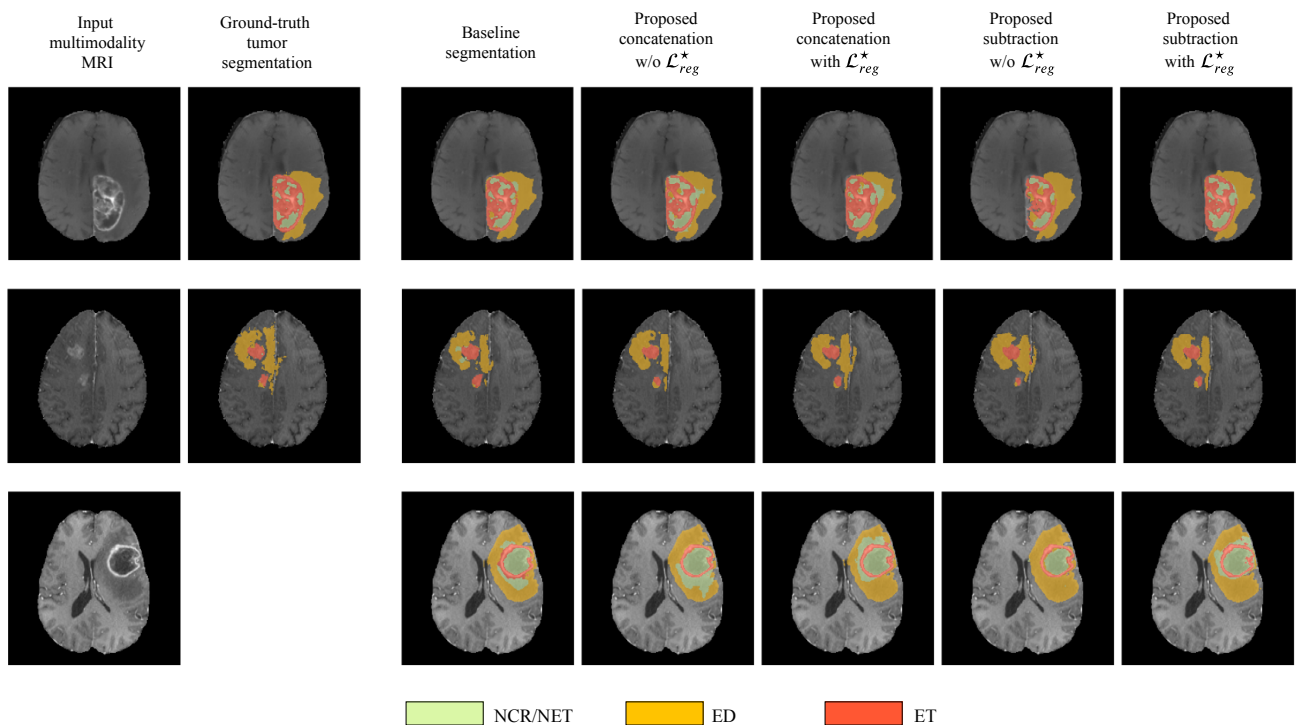


Figure 3. The segmentation maps produced by the different evaluated methods displayed on post-contrast Gadolinium T1-weighted modalities. We present the provided segmentation maps both on our test dataset and on the BraTS 2018 validation dataset. NCR/NET: necrotic core, ET: GD-enhancing tumor, ED: peritumoral edema.

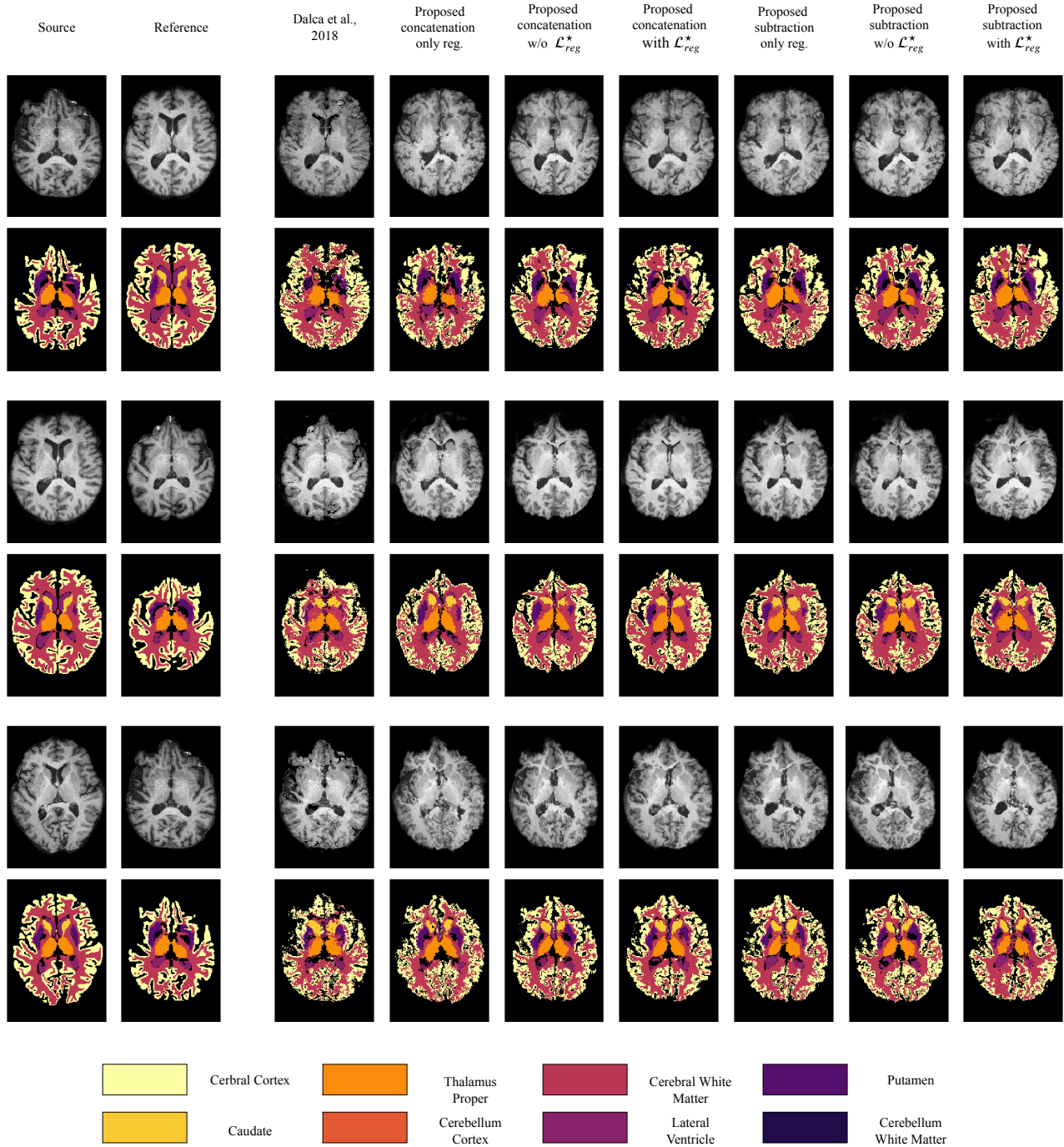


Figure 4. Qualitative evaluation of the registration performance for the different evaluated methods, displayed on T1 modalities. For an easier visualisation, we group left and right categories and only display the following 9 classes: caudate (Ca), cerebellum cortex (CbImC), cerebellum white matter (CbImWM), cerebral cortex (CebIC), cerebral white matter (CebIWM), lateral ventricle (LV), pallidum (Pa), putamen (Pu), ventral DC (VDC).

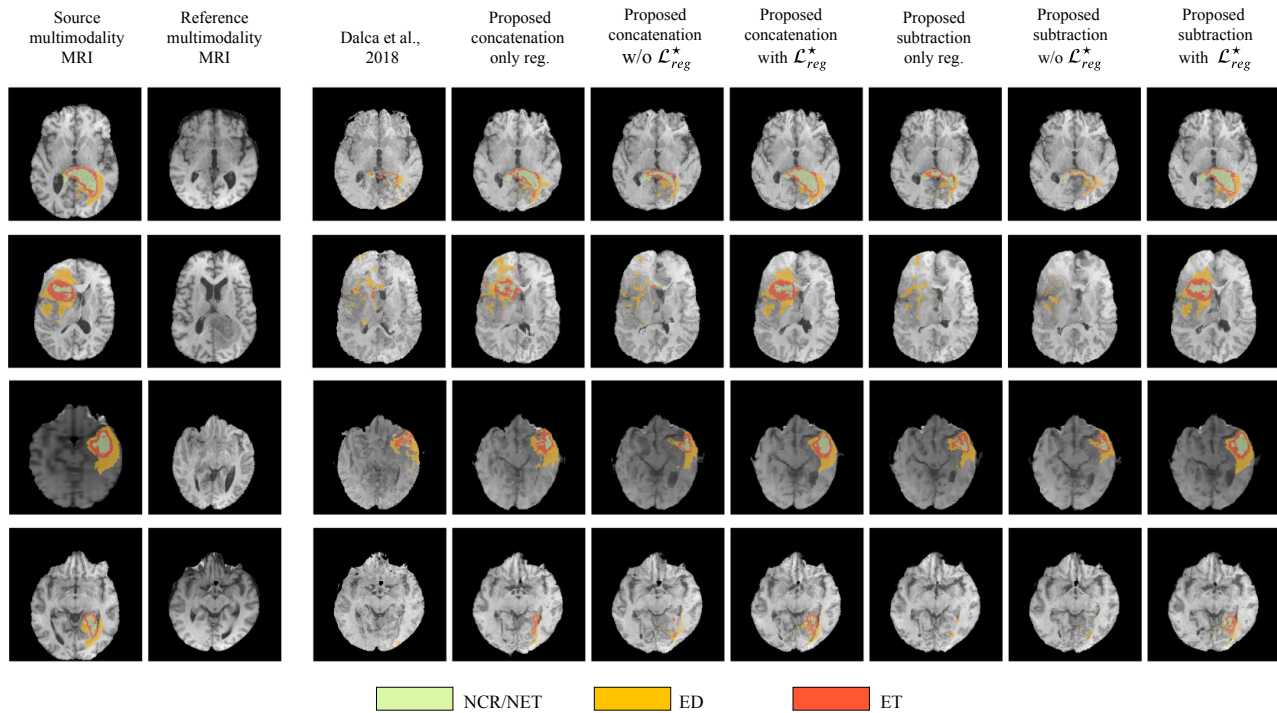


Figure 5. Qualitative evaluation of the tumor deformation of the different evaluated methods, displayed on T1 modalities. Each line is a sample, with source MRI in the first column to be registered on reference MRI in the second column. BraTS ground-truth annotations are plotted onto the source MRI. 7 models are benchmarked, one for each of the remaining columns which display the result of applying the predicted grid onto the source MRI. For each model and each line, the source ground-truth annotation masks of the source MRI were also registered with the predicted deformation grid, and the consequently obtained deformed ground-truth were plotted onto each deformed source MRI to illustrate the impact of all methods regarding the preservation of tumor extent.

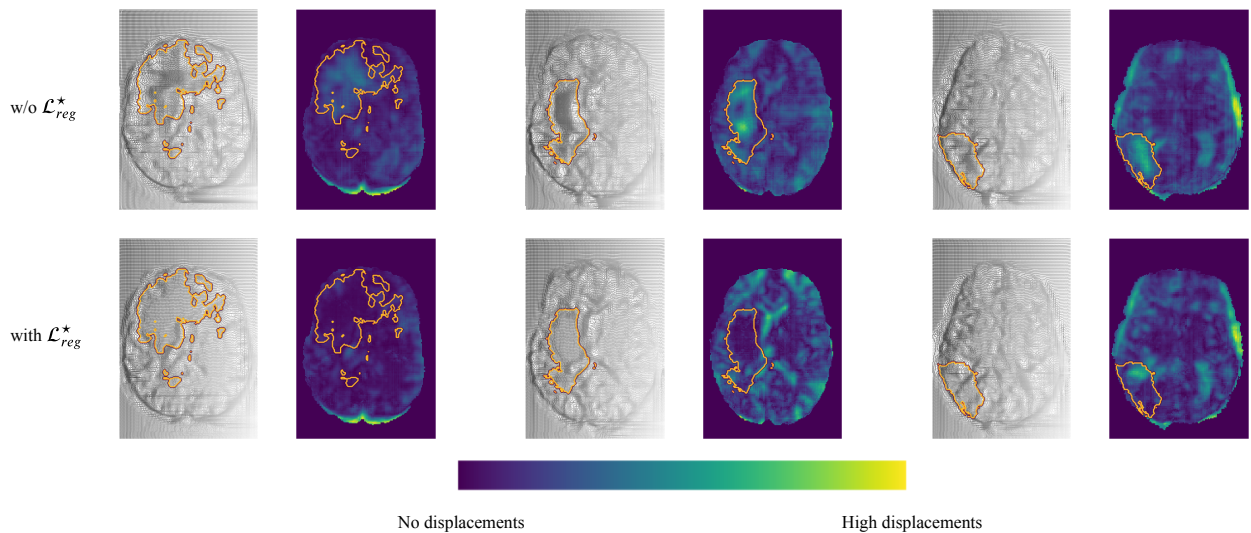


Figure 6. Comparison of the registration grid of the proposed model using the subtraction operation with and w/o \mathcal{L}_{reg}^* . This figure is obtained by sampling three random pairs of test patients, and computing the predicted registration fields, which are displayed by line for the two models, and in consecutive columns, one for each of the 3 dimensions, showing the registration field as a warped grid (grayscale) and as a colored map obtained by computing its norm pixelwise (blue-green map). Furthermore, the contour of the Whole Tumor is plotted on top of each image, obtained from the ground truth segmentation.