Unsupervised Multimodal Supervoxel Merging Towards Brain Tumor Segmentation

Guillaume Pelluet[1,2](✉) , Mira Rizkallah[1] , Oscar Acosta[3] ,
and Diana Mateus[1]

[1] Nantes Université, École Centrale Nantes, CNRS, LS2N, UMR 6004, F-44000
Nantes, France
[2] Hera-MI, SAS, Nantes, France
guillaume.pelluet@hera-mi.com
[3] Université de Rennes 1, LTSI, UMR, 1099 Rennes, France

Abstract. Automated brain tumor segmentation is challenging given the tumor's variability in size, shape, and image intensity. This paper focuses on the fusion of multimodal information coming from different Magnetic Resonance (MR) imaging sequences. We argue it is important to exploit all the modality complementarity to better segment and later determine the aggressiveness of tumors. However, simply concatenating the multimodal data as channels of a single image generates a high volume of redundant information. Therefore, we propose a supervoxel-based approach that regroups pixels sharing perceptually similar information across the different modalities to produce a single coherent oversegmentation. To further reduce redundant information while keeping meaningful borders, we include a variance constraint and a supervoxel merging step. Our experimental validation shows that the proposed merging strategy produces high-quality clustering results useful for brain tumor segmentation. Indeed, our method reaches an ASA score of 0.712 compared to 0.316 for the monomodal approach, indicating that the supervoxels accommodate well tumor boundaries. Our approach also improves by 11.5% the Global Score (GS), showing clusters effectively group pixels similar in intensity and texture.

Keywords: Brain tumor · Supervoxel · Merging · Graph · Clustering

1 Introduction

Identifying the edges of brain tumors and observing their evolution is critical to accurately assess disease progression and thus better guide the patient's treatment plan [9].

There is a multiplicity of brain imaging techniques, starting from the different Magnetic Resonance Imaging (MRI) sequences, providing complementary

information about brain tumors. However, multi-modality makes tumor segmentation, i.e., delineating the tumor's edges and quantifying the tumor's size, more complex. Commonly used sequences include T1, T2, FLAIR, and T1-weighted contrast-enhanced (T1CE). The visibility of glioma in the various sequences (modalities) is different. In the T1CE image, regions of the brain are similar to the tumor Edema region. In the T1CE, the active and necrotic regions of a tumor can be clearly distinguished. The intensities of edema and tumor regions are higher in the T2 sequence images and the FLAIR images, whereas the intensities of CerebroSpinal Fluid (CSF) are higher in the T2 and lower in FLAIR images. To sum up, one modality can present weak tumor edges but strong tumor features, while another may have strong edges but weak features. Many of the existing algorithms for brain tumor analysis focus on a single modality (e.g., a specific MRI sequence), limiting the available information to be exploited for segmentation.

Conversely, multimodal information can make the delineation and quantification more accurate, thanks to the modalities' complementarity. However, simultaneous processing different MRI sequences comprised of millions of voxels induces a significant increase in computational time. To tackle this problem, we propose to oversegment the original sequences with the idea to process supervoxels with similar information instead of the individual pixels. The concept of superpixel was originally introduced in [1] as a small homogeneous group of neighboring pixels. Hereafter, we refer to a supervoxel as an extension of a superpixel in the 3-D multi-modal setting.

We propose a two-stage unsupervised supervoxel-based approach. The first stage, performs an over-segmentation of the multimodal image with a supervoxel approach that approximates the boundaries of tumors and other objects in the multimodal image. The supervoxels are computed using an adaptation of the Scalable Simple Linear Iterative Clustering (SSLIC) algorithm [13]. Our adaption adds on a local regularity coefficient based on the variance [6] within the SSLIC algorithm. The coefficient increases the spatial constraint for supervoxels having high-intensity variances, and reduces it in areas with lower variances. Thereby, it allows supervoxel boundaries to capture perceptible objects with limited intensity variations. The second stage fuses multimodal supervoxels with a merging algorithm inspired by Fu *et al.* [5] to reduce the supervoxels' redundancy and their number prior to any classification task.

We evaluated our method on the publicly available multimodal BraTS 2020 dataset, which is a standard brain tumor segmentation benchmark [16]. Experiments show that the proposed merging produces highly accurate clusters compared to traditional monomodal approaches, thanks to the complementarity between modalities. We also demonstrate that using the local regularity coefficient allows generating more regular clusters on textures, better guiding the merging procedure. In the resulting segmentation after merging, the redundancy is reduced by a factor of 35 and the obtained supervoxels adhere very well to tumors boundaries.

2   Related Work

Brain tumor and lesion segmentation is often formulated as a pixel-wise semantic segmentation problem addressed with supervised learning approaches [4]. Among them, Convolutional Neural Networks (CNNs) have emerged as the current best-performing methods [15] taking different forms: 2D CCNs [2,18], 3D CNNs [3], or extended to Fully convolutional [12] or multimodal approaches [23]. Despite their good performance, pixel-wise methods suffer from high computational complexity due to the significant number of redundant pixels, particularly when dealing with multimodal images. This complexity affects both classical and learning-based algorithms. In the case of CNNs, multimodal images may require higher capacity networks, prone to overfitting if the training dataset is small. In this work, we take a step aside from pixel-wise semantic segmentation and focus on the unsupervised early fusion of multimodal information.

Compared to pixels, superpixels are more consistent with human visual cognition, contain less redundancy, and reduce noise. Superpixels generally allow to significantly improve the speed compared to pixel-based algorithms by analyzing pixels clusters [7]. These properties are useful for computationally expensive tasks, such as brain tumor segmentation in multi-sequence MRI images. Most superpixel-based algorithms cluster the image into a high number of redundant superpixels (called oversegmentation) by adding cuts to a graph or growing from predefined seeds [24]. Superpixel methods combined with conventional machine learning approaches have been used for brain tumor segmentation, demonstrating to be fast and robust to noise, initialization, and intensity non-uniformity [10,20]. However, these approaches neglect multimodal information in the superpixel step. Ignoring multimodality leads to of lack of adherence with weak boundaries, as noticed by Wang et al. [25]. Therefore, we opt for combining multimodal acquisitions, taking advantage of the complementary information to detect more detailed tumors structures and better adhere to borders.

Regarding other multimodal methods for brain tumor segmentation, Rahimpour *et al.* [19] compare early and late CNN fusion, favoring late fusion as it does not need an initial registration step. In our work, we opt for an early but unsupervised fusion which assumes pre-registered modalities. Soltaninejad *et al.* [22] also proposed an early multimodal fusion approach to produce supervoxel boundaries across multiple MR sequences, enforcing adherence to weak structures boundaries. However, similar to the monomodal case, the algorithm results in a large number of redundant superpixels, which unnecessarily increases computation time and can lead to a higher false-positive rate. For this reason, we propose two contributions to reduce supervoxels redundancy in the multimodal case. First, a variance constraint inspired by the work of Giraud *et al.* [6], proposed in the context of natural images to better account for textured regions; and second, a supervoxel merging step.

Outside the brain tumor segmentation literature, there has been interest in superpixel and supervoxel merging approaches. Luengo *et al.* [14] proposed a method that achieves high segmentation performance while reducing the number of redundant superpixels in the image, based on an iterative splitting and merg-

ing algorithm. Focusing on the scale, Fu *et al.* [5] introduced a multiscale approach for superpixel merging in the RGB color space. The method uses multiple features to calculate a dissimilarity score between pairs of superpixels, including color, texture, and common border length. Moreover, it simplifies the merging graph to accelerate the merging procedure. For these two reasons, which are relevant in our multimodal MRI case, we rely on Fu's multiscale approach for supervoxel merging. Our experimental validation shows qualitatively and quantitatively the pertinence of our two contributions: the variance constraint and the merging approach. Our approach combining multimodal supervoxels, the variance constraint, and the merging step, improves tumor boundary adherence and significantly reduces supervoxel redundancy.

3   Methods

Let multiple images of the same anatomy be acquired with different modalities and then registered to form the multimodal image $\mathbb{I} = [I_1, I_2, \ldots, I_M]$. $\mathbb{I}$ is a 3-D volume whose every voxel contains an M-dimensional vector. Our goal is to find a single partition S of non-overlapping supervoxels S_i, such that, $S = \bigcup_{i=1}^{n} S_i$ taking into account intensities and borders in all modalities. To this end, we propose a two-steps method. First, an initial oversegmentation is performed with the SSLIC algorithm [13], refined with a variance constraint to better model the texture. As a result we obtain an initial supervoxel clustering (See Sect. 3.1). However, the oversegmentation can lead to a substantial number of supervoxels even for a small tumor. This creates a burden for later tasks, such as classification. To reduce the final number of supervoxels, a second step is necessary. Inspired by the work of Fu *et al.* [5], we construct a graph $\mathcal{G}$ over the oversegmentation and merge similar vertices to obtain a more meaningful segmentation (See Sect. 3.2).

3.1   Oversegmentation Based on Supervoxels

Supervoxels are irregular image blocks composed of adjacent voxels with similar texture, intensity, and brightness features. Currently, there are two common types of supervoxel segmentation algorithms. The first one is based on graph theory and the second on Gradient Ascent. To the later category belongs the well-known Simple Linear Iterative Clustering (SLIC) approach [1] and its ITK version [8]. We rely on SSLIC with multimodal features [13] to obtain a first oversegmentation of the image. By multimodal features we mean that each voxel is characterized by an M-dimensional vector containing the intensities for that pixel across all modalities. First, an initial clustering is given and then the clustering is improved iteratively until convergence (refer to [13] for details).

We propose an adaption of the SSLIC algorithm ($SSLIC_{Var}$), that modulates the supervoxel compactness according to the supervoxels feature variance. Initially introduced by Giraud *et al.* [6] in the context of natural imaging in 2D, we bring this constraint to the medical image analysis field, extending it

for the M-dimensional case. The standard SSLIC framework [13] only requires the number of superpixels and a single parameter m. In our adapted version, each supervoxel S_i has a different parameter m_i setting its shape regularity (i.e. compactness). This parameter is computed according to the mean feature (luminance in our case) variance per supervoxel across modalities:

$$m_i = m * \exp\left(\overline{\frac{\sigma_i^2(F_{\mathrm{mod}})}{\epsilon}}\right) \tag{1}$$

where $\sigma_i^2(F_{\mathrm{mod}})$ is the luminance variance within the supervoxel S_i in a modality, $\bar{\cdot}$ is the mean operator and ϵ is a scaling parameter. At the output of this step, we have an oversegmentation of our 3D multimodal volume $\mathbb{I}$.

3.2   Supervoxels Merging

The oversegmentation produced by the supervoxel-based method already reduces some redundant information. However, the SSLIC approach is sensitive to the seeds initialization, which constraints the final number of clusters. Flat objects in the image, such as tumors exhibiting low texture and small intensity variation, are still composed of redundant supervoxels. With the aim of further reducing the redundancy, we use a method inspired from the work of Fu *et al.* [5] and apply it in the context of multimodal MRI. The oversegmentation is transformed into a Region Adjacency Graph (RAG) $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, with the set of vertices $\mathcal{V} = \{v_1, v_2, ..., v_n\}$ and n the number of supervoxels. Edges $\mathcal{E}$ represent connections between adjacent supervoxels and their weights denote the dissimilarity based on the intensity and texture features. The dissimilarity of two supervoxels i and j, named $w_{i,j}$, is defined as Eq. 2.

$$w_{i,j} = \exp\left(-\frac{(\frac{\alpha \cdot D_c(i,j) + \beta \cdot D_t(i,j)}{\alpha+\beta})^2}{\gamma}\right), \tag{2}$$

where $D_c(i,j)$ and $D_t(i,j)$ are the intensity and texture dissimilarities, α and β their respective adjustable weights, and γ governs how close to each other features are. More specifically,

$$D_c(i,j) = \sqrt{\sum_{\mathrm{mod}=1}^{M} \Delta Y_{\mathrm{mod}}(i,j)}, \tag{3}$$

where $\Delta Y_{\mathrm{mod}}(i,j) = (Y_{\mathrm{mod}}^i - Y_{\mathrm{mod}}^j)^2$ and $Y_{\mathrm{mod}}^i, Y_{\mathrm{mod}}^j$ are the average luminance values in the ith and j^{th} supervoxels respectively. $D_t(i,j)$ is the texture dissimilarity computed in [5] as :

$$D_t(i,j) = \sqrt{\sum_{\mathrm{mod}=1}^{M} \Delta H_{\mathrm{mod}}(i,j)}, \tag{4}$$

where $\Delta H_{\mathrm{mod}}(i,j)$ is the Manhattan distance between the histograms of supervoxels i and j as in [5]. The distance measures were normalized in a range of $[0;1]$ to be efficiently combined. Some brain tissues, as is the case of tumors, have lower textures and high intensity, which can result in an imbalance between intensity and texture features. Because of these complex cases, the adjustable weights from Eq. 2 were manually adjusted to better split the dissimilarity between normal and tumor tissues as defined in Sect. 4.3. Once the dissimilarity measures over supervoxels and graph weights are computed, the supervoxel merging algorithm takes place to reduce information redundancy and achieve finer clustering. However, the Region Adjacency Graph (RAG) connects each supervoxel to all its neighbors. So, it is very computationally expensive to directly start merging the nodes with high similarity since the number of edges and nodes is still too large. To accelerate the merging process, a Nearest Neighbor Graph (NNG) [17] is determined based on the RAG. The NNG efficiently determines paired supervoxels that are the most similar. Here, the NNG is calculated using the Kruskal algorithm [11], which significantly reduces the number of edges and overall the search space, allowing for a more computationally efficient merging. The merging algorithm is iteratively computed until no edges in the NNG have weights inferior to a given threshold $\mathcal{T}$ which is defined as in Eq. 5:

$$\mathcal{T} = \frac{\sum_j(\min e_j - \sigma(e_j))}{n},\tag{5}$$

with e_j one of the edges connected to supervoxel i, that is, $e_j \in \{w_{ij}\}$, $j \in \mathcal{N}_i$ and σ denotes the standard deviation.

4   Experiments

4.1   Experimental Setup

Experiments are performed on the publicly available multimodal BraTS 2020 dataset, which is a standard brain tumor segmentation benchmark [16]. The dataset is composed of real brain MRI exams including T1, T1CE, T2, and FLAIR sequences, acquired from 19 institutions for 369 subjects. The ground truth is provided for each exam in form of contours manually delineated by experts. Three tumor subregions were annotated: contrast-enhancing tumor, non-enhancing/necrosis combined, and edema. Images are 3D volumes with a size of $[155 \times 240 \times 240]$ (DxWxH) and an isotropic resolution of 1 mm. The sequences from the dataset are co-registered to the same anatomical shape and skull-stripped by the BraTS maintainer. Images are cropped to remove the background area at the edges and normalized independently for each modality between $[0;1]$.

4.2   Quality Assessment Methods

We use several reference (using ground-truth) and no-reference segmentation assessment metrics to evaluate the performance of the proposed unsupervised

segmentation method in delineating tumor tissues and keeping meaningful voxels disparities. The Achievable Segmentation Accuracy (ASA) score is computed in the tumor's region to assess the accuracy of the supervoxels boundaries with respect to the ground truth. The wVar and Moran' Index (MI) quantify respectively the disparity within and between clusters. More precisely, the wVar assesses the luminance disparity of within each cluster, while MI is a spatial autocorrelation measure characterizing the degree of similarity among supervoxels. Since the SSLIC oversegmentation is highly redundant, MI is an effective measure to show the advantage of the merging approach. The best value for wVar and MI is 0 which indicates the absence of redundancy. The Global Score (GS) is defined as the average of wVar and MI and is used as a final metric with ASA. We also use the number of supervoxels in the image (Supervoxel count) to quantify the improvement brought by the merging algorithm. For the no-reference metrics, in the monomodal setting, the final results are computed as an average through all modalities for all subjects. In the multimodal setting, the final results correspond to the average across all subjects. Since the wVar et MI scores provide one measure per modality, we keep the minimal value for each supervoxel across modalites. The evaluation is done in this way to put forward the discriminative power of the different modalities. The other scores (ASA and count) directly provide a single measurement per subject.

4.3   Implementation Details

SSLIC and merging algorithms are dependent on input parameters. The quality of the output clustering with SSLIC depends on the parameters K and m. K is the number of supervoxels, which in our case is defined as the smallest desired isotropic supervoxel size $K = [10, 10, 10]$. As multimodal images are normalized independently between $[0, 1]$, the compactness factor m is defined at 0.1. This value better balances intensity and spatial features as spatial features are not normalized to the range $[0, 1]$. The variance parameter ϵ used to balance the influence of the variance on the local compactness is set to 0.01. The hyperparameters α, β, γ have been empirically defined at 0.5, 0.1, and 0.1 to balance feature importance. Several orders of values have been tested to retain the parameter set with higher ASA. The parameters used in the histogram texture similarity are set to 32 for the number of bins, 8 for the number of angles, and 10 for the histogram bin size. The whole process takes around 40s for an image of shape $[4 \times 155 \times 240 \times 240]$ with the first axis corresponding to the number of modalities M. The SSLIC algorithm and the feature extraction were computed on 12 threads with 32 GB of memory.

4.4   Experimental Results

In our experiments, we assess the benefit of exploiting multimodal information in computing supervoxels, the effectiveness of including variance as a regularity coefficient in the SSLIC and the impact of the merging algorithm relying on
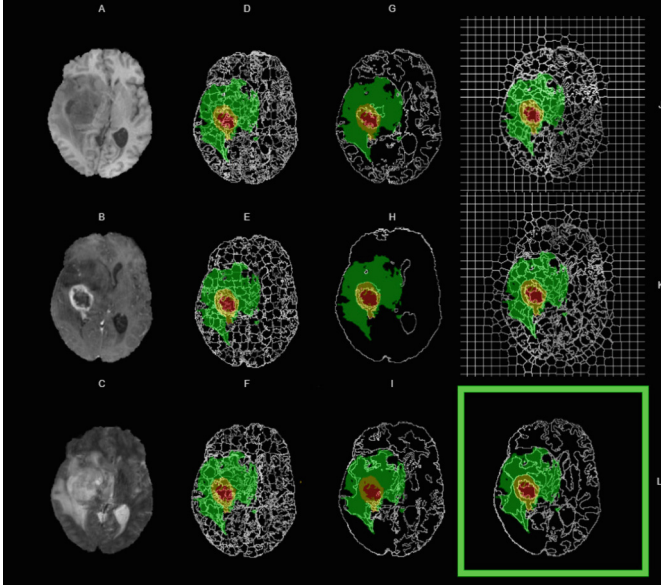
Fig. 1. The first column is an axial cross-section over 3 MRI sequences: T1 (A), T1_CE (B), T2 (C). The second column (D, E, and F) are the supervoxels computed using $Mono\_SSLIC$ on the 3 modalities independently. The third column corresponds to the result of the merging procedure applied on the previously computed supervoxels on each modality ($Mono\_SSLIC_{Merged}$). In the last column, J corresponds to the resulting segmentation of $Multi\_SSLIC$ computed on the 3D volume $\mathbb{I}$ composed of the different modalities, K is the result of SSLIC computed on $\mathbb{I}$ with the local regularity coefficient ($Multi\_SSLIC_{Var}$) and L is the proposed method including multimodal SSLIC followed by the merging procedure with the local regularity coefficient ($Multi\_SSLIC_{Var\_Merged}$). The ground-truth overlay is represented by green, red, and yellow (Edema, necrosis, and active tumor). (Color figure online)
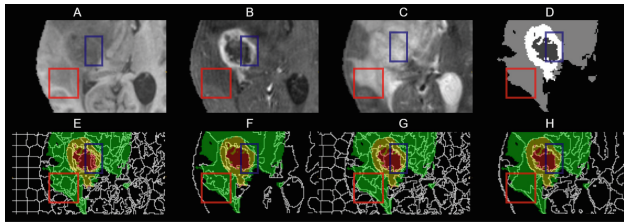
Fig. 2. (TOP) Original multimodal images zoomed in around the tumor region. Modalities are T1 (A), T1CE (B), T2 (C), and the ground truth (D). (Bottom) $Multi\_SSLIC$, $Multi\_SSLIC_{Merged}$, $Multi\_SSLIC_{Var}$ and $Multi\_SSLIC_{Var\_Merged}$ (E-H). Blue and red squares show local adaptive regularity influence on supervoxel homogeneity and compactness. The ground-truth overlay is represented by green, red, and yellow (Edema, necrosis, and active tumor). (Color figure online)

Table 1. Performance measurements computed with our own implementation of the scores added to superpixel benchmark [24]

Monomodal method	ASA	wVar	MI	GS	Supervoxel count
$Mono\_SSLIC$	.625	**.314**	.398	.356	8629.770
$Mono\_SSLIC_{Var}$	**.676**	.346	.362	.354	8233.930
$Mono\_SSLIC_{Merged}$	.290	**.314**	**.220**	**.267**	204.625
$Mono\_SSLIC_{Var\_Merged}$	.316	.348	.318	.333	203.904

Multimodal method	ASA	wVar	MI	GS	Supervoxel count
$Multi\_SSLIC$	.687	.493	.417	.455	8626.370
$Multi\_SSLIC_{Var}$	.648	.505	.383	.444	8163.21
$Multi\_SSLIC_{Merged}$	.673	.483	**.337**	.409	301.417
$Multi\_SSLIC_{Var\_Merged}$	**.712**	**.458**	.349	**.403**	298.42

colors and textures features on the segmentation accuracy. To this end, we compare 4 unsupervised segmentation methods applied in both the monomodal and the multimodal settings: SSLIC applied without ($SSLIC$) or with ($SSLIC_{Var}$) the adaptive local variance regularity coefficient, SSLIC followed by the merging step without ($SSLIC_{Merged}$) or with the adaptive local variance regularity coefficient ($SSLIC_{Var\_Merged}$, ours). The former methods are applied both in monomodal ($Mono$) and multimodal ($Multi$) settings.

Figures 1 and 2 show some qualitative results of applying the 4 segmentation methods to one subject with 4 modalities FLAIR, T1, T1CE, and T2. To further illustrate the performance of the proposed approaches, we report in Table 1 several quality metrics computed on the segmentations obtained in both the monomodal and multimodal settings.

The Benefit of Multimodality. As depicted in Fig. 1 J, applying the segmentation on multimodal images successfully takes into account the heterogeneous information from different modalities to cluster the image. On the contrary, in Fig. 1 D–F (results generated from $Mono\_SSLIC$), the clusters do not adhere completely to the ground truth tumor boundaries on the T1 and T2 modalities, since the complete information concerning the tumor is not fully present and multimodal information can not be efficiently exploited. In Fig. 1 G–I, we show the results of the merging applied independently on the three modalities with ground-truth overlay. It is clear that the T2 modality gives more information about Edema tissue whereas T1CE further characterizes the tumor's tissue. A more accurate clustering of the tumor can be seen in Fig. 1 J–L.

As shown in Table 1, the multimodal approaches i.e. $Multi\_SSLIC_{Var}$ and $Multi\_SSLIC_{Var\_Merged}$ perform better in terms of ASA compared to the monomodal approaches. Multimodal clustering exploits all the available information from different modalities and produces an accurate segmentation. We found that the best performing approach is the $Multi\_SSLIC_{Var\_Merged}$ which improves the clustering accuracy by 5.2% for the ASA Score and 25% for the GS with multimodal information. Indeed, all modalities give different complemen-

tary information about tissues. Thereby, using all available information to merge supervoxels while keeping important tissue properties, such as tumors texture, improves qualitative results as well as ASA, and GS scores.

Impact of Locally Adapting the Superpixel Regularity. Including variance inside the SSLIC algorithm allows to automatically adapt the regularity coefficient to highly textured supervoxel s and high-intensity supervoxels without manually adapting m. This makes the supervoxels more homogeneous as well as more compact, resulting in a better final clustering accuracy as shown in the Fig. 2. Blue and red squares in Fig. 2 F and H ($Multi\_SSLIC_{Merged}$ and $Multi\_SSLIC_{Var\_Merged}$), show the influence of using the local regularity coefficient on the compactness of the merged supervoxels. The resulting supervoxels are more compact and differ from their neighbors. We can see in the red square of Fig. 2 H that supervoxels have been correctly computed with more compactness and have been merged into a bigger supervoxel. Furthermore, from the quantitative results in Table 1, we can see that the local adaptive regularity coefficient $*_{Var}$ improves the results in terms of accuracy (ASA) and GS for the methods applied in both monomodal and multimodal settings (excepts for $Multi\_SSLIC$ and $Multi\_SSLIC_{Var}$). The variance of the supervoxel is an important factor to take into account in the segmentation algorithm. The MI is almost the same for both $Multi\_SSLIC_{Merged}$ and $Multi\_SSLIC_{Var\_Merged}$ demonstrating the robustness of the merging step to variance's disparity across supervoxels.

Performance of the Merging Algorithm. In the monomodal setting, in a modality where tumor tissues are not distinct, merging similar neighboring supervoxels reduces the tumor boundary accuracy. For example, in Fig. 1 H, supervoxels computed independently on the T1CE modality are not accurately merged since this modality highlights only the active tumor while other tumor tissues are not visible. This results in a poor ASA score for T1CE, therefore penalizing the final average ASA score. As such, computing the average ASA across modalities highlights the lack of the multimodal discriminant power (making use of visible tumors parts in all modalities). The merging approach applied in the multimodal setting is capable of reducing the number of supervoxels by a factor of 35 (column "Supervoxel count" in Table 1) and decreasing the redundancy (MI) by 0.21% in average compared to the initial oversegmentation). The texture homogeneity inside the merged supervoxels has been kept which demonstrates that our algorithm merges similar supervoxels. It is also interesting to note the wVar obtained on the results of applying $Mono\_SSLIC$ or $Mono\_SSLIC_{Merged}$ is approximately similar. This can be explained by the fact that the clustering was initially correct for the $Mono\_SSLIC$ step without merging.

5   Conclusion

In this work, we proposed a novel approach of merging supervoxels in a multimodal setting towards brain tumor classification. We showed that our methods

applied on multimodal images are capable of exploiting the complementarity between different modalities producing very accurate clusters compared to traditional monomodal approaches. Our approach $Multi\_SSLIC_{Var\_Merged}$ improved the clustering accuracy by 5.2% for the ASA Score and 25% for the GS. The redundancy of supervoxels is also reduced by a factor of 35, decreasing the computational time, and making the resulting oversegmentation more suitable to be combined with a neural network classifier. Several open questions remain to be tackled in a future work. First, one drawback of the proposed approach is its dependency on prior registration of multiple modalities. Bipartite Graph Matching [21] seems to be an efficient way to alleviate this constraint. Moreover, taking into account radiomics and deep features in the computation of the supervoxels could also improve the adherence of initial over-segmentation or merged supervoxels to contrasted tissues, therefore resulting in more homogeneous final clustering.

References

1. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: Slic superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intell. **34**(11), 2274–2282 (2012). https://doi.org/10.1109/TPAMI.2012.120
2. Ahuja, S., Panigrahi, B.K., Gandhi, T.: Transfer learning based brain tumor detection and segmentation using superpixel technique. In: 2020 International Conference on Contemporary Computing and Applications, IC3A 2020, pp. 244–249 (2020). https://doi.org/10.1109/IC3A48958.2020.233306
3. Ali, M., Gilani, S.O., Waris, A., Zafar, K., Jamil, M.: Brain tumour image segmentation using deep networks. IEEE Access **8**, 153589–153598 (2020). https://doi.org/10.1109/ACCESS.2020.3018160
4. El-Melegy, M.T., El-Magd, K.M.A., Ali, S.A., Hussain, K.F., Mahdy, Y.B.: A comparative study of classification methods for automatic multimodal brain tumor segmentation. In: 2018 International Conference on Innovative Trends in Computer Engineering (ITCE). pp, 36–41 (2018). https://doi.org/10.1109/ITCE.2018.8316597
5. Fu, Z., Sun, Y., Fan, L., Han, Y.: Multiscale and multifeature segmentation of high-spatial resolution remote sensing images using superpixels with mutual optimal strategy. Remote Sens. **10**(8) (2018). https://doi.org/10.3390/rs10081289
6. Giraud, R., Ta, V., Papadakis, N., Berthoumieu, Y.: Texture-aware superpixel segmentation. In: IEEE International Conference on Image Processing (ICIP), abs/1901.11111 (2019)
7. Ibrahim, A., El-Kenawy, E.S.M.: Image segmentation methods based on superpixel techniques: a survey. J. Comput. Sci. Inf. Syst. **2020**(6), 1–10 (2020)
8. Johnson, B., Xie, Z.: Unsupervised image segmentation evaluation and refinement using a multi-scale approach. ISPRS J. Photogramm. Remote Sens. **66**(4), 473–483 (2011). https://doi.org/10.1016/j.isprsjprs.2011.02.006

9. Johnson, D.R., et al.: Glioma response assessment: classic pitfalls, novel confounders, and emerging imaging tools. Br. J. Radiol. (2019)
10. Khosravanian, A., Rahmanimanesh, M., Keshavarzi, P., Mozaffari, S.: Fast level set method for glioma brain tumor segmentation based on superpixel fuzzy clustering and lattice Boltzmann method. Comput. Methods Programs Biomed. **198**, 105809 (2021). https://doi.org/10.1016/j.cmpb.2020.105809
11. Kruskal, J.B.: On the shortest spanning subtree of a graph and the traveling salesman problem. Proc. Am. Math. Soc. **7**(1), 48–50 (1956)
12. Liqiang, Y., Erdt, M., Lipo, W.: Adaptive transfer learning to enhance domain transfer in brain tumor segmentation. In: Proceedings - International Symposium on Biomedical Imaging 2021-April, pp. 1873–1877 (2021). https://doi.org/10.1109/ISBI48211.2021.9434100
13. Lowekamp, B.C., Chen, D.T., Yaniv, Z., Yoo, T.S.: Scalable simple linear iterative clustering (SSLIC) using a generic and parallel approach. Kitware, Inc. (2018)
14. Luengo, I., Basham, M., French, A.P.: SMURFS: superpixels from multi-scale refinement of super-regions. In: British Machine Vision Conference 2016, BMVC 2016 (2016). https://doi.org/10.5244/C.30.4
15. Magadza, T., Viriri, S.: Deep learning for brain tumor segmentation: a survey of state-of-the-art. J. Imaging **7**(2) (2021). https://doi.org/10.3390/jimaging7020019
16. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Med. Imaging **34**(10), 1993–2024 (2015). https://doi.org/10.1109/TMI.2014.2377694
17. Paterson, M.S., Yao, F.F.: On nearest-neighbor graphs. In: Kuich, W. (ed.) ICALP 1992. LNCS, vol. 623, pp. 416–426. Springer, Heidelberg (1992). https://doi.org/10.1007/3-540-55719-9_93
18. Pereira, S., Pinto, A., Alves, V., Silva, C.A.: Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Trans. Med. Imaging **35**(5), 1240–1251 (2016). https://doi.org/10.1109/TMI.2016.2538465
19. Rahimpour, M., Goffin, K., Koole, M.: Convolutional neural networks for brain tumor segmentation using different sets of MRI sequences. In: 2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), pp. 1–3 (2019). https://doi.org/10.1109/NSS/MIC42101.2019.9059769
20. Rehman, Z.U., Naqvi, S.S., Khan, T.M., Khan, M.A., Bashir, T.: Fully automated multi-parametric brain tumour segmentation using superpixel based classification. Expert Syst. Appl. **118**, 598–613 (2019). https://doi.org/10.1016/j.eswa.2018.10.040
21. Serratosa, F.: Fast computation of bipartite graph matching. Pattern Recogn. Lett. **45**, 244–250 (2014). https://doi.org/10.1016/j.patrec.2014.04.015
22. Soltaninejad, M., et al.: Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in FLAIR MRI. Int. J. Comput. Assist. Radiol. Surg. **12**(2), 183–203 (2016). https://doi.org/10.1007/s11548-016-1483-3
23. Soltaninejad, M., et al.: Supervised learning based multimodal MRI brain tumour segmentation using texture features from supervoxels. Comput. Methods Programs Biomed. **157**, 69–84 (2018). https://doi.org/10.1016/j.cmpb.2018.01.003
24. Stutz, D., Hermans, A., Leibe, B.: Superpixels: an evaluation of the state-of-the-art. Comput. Vis. Image Underst. **166**, 1–27 (2018). https://doi.org/10.1016/j.cviu.2017.03.007
25. Wang, W., Chen, C., Ding, M., Li, J., Yu, H., Zha, S.: TransBTS: multimodal brain tumor segmentation using transformer (2021)