Spherical coordinates transformation pre-processing in Deep Convolutional Neural Networks for brain tumor segmentation in MRI

Carlo Russo1 · Sidong Liu1,2 · Antonio Di Ieva1

Received: 10 April 2021 / Accepted: 21 October 2021 / Published online: 2 November 2021
© International Federation for Medical and Biological Engineering 2021

Abstract
Magnetic Resonance Imaging (MRI) is used in everyday clinical practice to assess brain tumors. Deep Convolutional Neural Networks (DCNN) have recently shown very promising results in brain tumor segmentation tasks; however, DCNN models fail the task when applied to volumes that are different from the training dataset. One of the reasons is due to the lack of data standardization to adjust for different models and MR machines. In this work, a 3D spherical coordinates transform during the pre-processing phase has been hypothesized to improve DCNN models’ accuracy and to allow more generalizable results even when the model is trained on small and heterogeneous datasets and translated into different domains. Indeed, the spherical coordinate system avoids several standardization issues since it works independently of resolution and imaging settings. The model trained on spherical transform pre-processed inputs resulted in superior performance over the Cartesian-input trained model on predicting gliomas’ segmentation on Tumor Core and Enhancing Tumor classes, achieving a further improvement in accuracy by merging the two models together. The proposed model is not resolution-dependent, thus improving segmentation accuracy and theoretically solving some transfer learning problems related to the domain shifting, at least in terms of image resolution in the datasets.

Keywords Brain tumor · Deep Convolutional Neural Network · MRI segmentation · Spherical coordinates

1 Introduction
In medical imaging, computer vision research, and many other different domains, Deep Learning (DL) algorithms have achieved impressive results, but only in fields where a large annotated dataset was previously available [1]. One of the major limitations of DL models is the lack of generalization when a model trained on a restricted dataset is applied to data from other institutions, introducing unpredictable results in the segmentation/classification accuracy [1]. The failure of actual models on MRI sequences is partially attributable to the lack of standardization of input images between different domains and the lack of availability of big datasets to train the DL model on a new domain. An exemplary application of a DL model is the automatic recognition and segmentation of brain tumors, such as gliomas. Since 2012, the Multimodal Brain Tumor Segmentation (BraTS) challenge is organized yearly within the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) (http://brain-tumor-segmentation.org/) [2], with the main aim of identifying the best model for automatic segmentation of MR images (MRI) of high-grade gliomas (HGG) and low-grade gliomas (LGG). The segmentation task aims to automatically recognize the total area containing the tumor (called Whole Tumor, WT, region), as well as the Necrosis and Active tumor cells area (Tumor Core, TC, and Enhancing Tumor, ET, areas, the latter contained in TC area). The DL models submitted to the BraTS challenge are achieving good results when trained on a relatively small dataset, consisting of about 400 cases of HGGs and LGGs. Despite using image augmentation techniques on such a dataset, the average Dice scores of the best methods are reaching less than 90% accuracy on the WT class (average 0.88 and median 0.92), and around 80% on the ET and TC classes (average 0.78 and median 0.85 in ET, average 0.82 and median 0.91 in TC) [3]. Moreover, in spite of the heterogeneous source of the images within the BraTS
dataset, there are still doubts about the translation of such models to different datasets from other institutions: indeed, some biases, such as distributional shift, are still not fully addressed by current AI methodologies [4, 5].

The lack of large brain tumor imaging labelled datasets and the need to overcome the heterogeneity of the datasets, as well as the need to mimic human vision and experts’ interpretation, in particular neuroradiologists’ and neurosurgeons’ expertise to identify and segment brain tumors, have triggered our current research in regard to new methods for pre-processing data representation, i.e., before feeding DL models. A new method to overcome these issues has been proposed here, by attempting to emulate the cognitive workflow that an expert in neuroradiology (e.g., a trained neuroradiologist or a neurosurgeon) would perform for tumor recognition and segmentation.

In the present paper, we briefly summarize other approaches in Sect. 2. Details of the transformation used in our novel method are described in Sect. 3, while Sect. 4 describes a test case/real application example of the method and the corresponding results. Sections 5 and 6 summarize advantages and limits of the method.

2 Background and related works

2.1 Background

In everyday clinical practice, medical experts’ eyes track scans on each single image slice, scouting for particular features in targeted anatomical regions [6, 7]. During this phase, image resolution, size, and orientation of the object, among others, are not salient features of the cognitive analysis. For such a reason, DL models with strong image augmentation based on rotation and zooming usually work better than others [8]. Furthermore, radiologists in general do not look at the entire image, rather their eyes follow an apparent path and focus on one region at a time [9–14]. Our hypothesis is that DL models’ accuracies can be improved by changing the representation of the image, e.g., transforming the standard Cartesian coordinates into polar-based coordinates, and the corresponding 3D spherical coordinate system for 3D volumes in the pre-processing phase. The transformation introduces a new invariant constraint compared to the classic Convolutional Neural Networks that are natively invariant to translation. Thus, the transformed object can be analyzed in a rotational/scale invariance. Furthermore, the new object will not be “position invariant” anymore, since each transformation is dependent upon the choice of an origin point. The workflow of progressive refining of the origin points during the observation process has been replicated in the steps of the new proposed method, as described below.

Polar and spherical representation examples of segmented radiologic images of a brain tumor (a glioblastoma in the right hemisphere) are shown in Fig. 1.

A polar/spherical coordinate system pre-processing was used to feed the CNN models as it may offer the following advantages:

1) The representation of the volume in polar and spherical coordinates enables the CNN to be invariant to rotation, since two volumes rotated differently have the same spherical coordinate system representation but just shifted in the angles’ planes by a distance equivalent to the rotation angle (Fig. 2C and 2D). A scaled object has the same polar representation scaled on the third direction, i.e., the distance radius, by a zooming factor equivalent to the one applied in the scaling process. This scaled representation can be avoided when considering the maximum radius to be the farthest distance from the origin to the edges of the image (e.g., the brain’s surface or just the image’s edge in a different resolution dataset) (Fig. 2E and 2F).

Such a process enables the CNN to learn scale invariant filters from spherical representation of volumes, that will be also invariant, or at least less susceptible, to different image resolutions within the original dataset and to different location and shapes of the tumor. When considering dataset of different resolutions, the original resolution is preserved better compared to downsampling methods (e.g., the 1 mm isotropic sampling used in BraTS challenges), since mapping the volume to the spherical transformation preserves information details around the origin point. Classic CNN models are invariant to region shifting; therefore, by this way, CNN filters become invariant to rotation, scaling, and resolution (at least within a specific scaling and resolution range), although this will depend on the choice of the origin point.

2) The spherical system represents one region at a time. Since the spherical representation needs to be centered according to a chosen origin, the DL model will scout a single region at a time, focusing more on the centered region and minimizing the weighting of the most peripheral ones. This is also a limit of the technique but it is in fact closer to the human vision system, whereby human experts’ eyes are focused on salient regions rather than on their peripheral features [6, 12–15]. The ensemble of the regions will then give the fully segmented volume. It is worth noting that this is different from the way some CNNs works on patches or subregion of the original volume: in the proposed method, models that takes the full volume as an input are preferred, and spherical transformation includes details of regions around the origin point as well as distant regions. Particularly, spherical transformed volumes contain voxels at high resolution within the target area as well as surrounding voxels at
lower resolution far from the target area. Patches contained in this representation are not scale-dependent, but can appear in a different way during the training phase when the origin points are randomly chosen.

3) Spherical representation will enable the analysis of the growing surface of a centered object to be improved. The tumor biologically expands from single points growing on the border in a regular or irregular way[16]. Once the tumor segmentation is performed, this representation can be used to enable better classification of the tumor type and grade just by evaluating the irregularity of the 3D spherical surface[17], avoiding issues caused by different orientation and zooming of objects, or for any therapeutic purposes (e.g., planning of radiotherapy/radiosurgery on the targeted region, intraoperative navigation for surgery). The tumor’s irregularity and complexity can be evaluated by means of radiomic parameters [18–20] and/or computational fractal-based analysis (e.g., computation of the fractal dimension), as investigated in other studies[21–28].

2.2 Related works

To the best of our knowledge, our study is the first attempt to use polar and spherical coordinates transform to improve image pre-processing for DL models. There have been some
studies which aim to induce the rotation and scale invariance to the DCNN. For example, Harmonic Networks (H-Nets) have been proposed by Worrall et al. to improve rotation invariance [29]. In this work, DCNN has been transformed to use harmonic filters instead of regular CNN filters, thus keeping the network parameter efficient without the necessity of including additional layers and filters for performing rotational invariance. H-Nets can be implemented in current DL frameworks with minor engineering. H-Nets are currently applicable only on 2D images, although the authors proposed to expand their work to analyze 3D volumes as well [29].

Depeursinge et al. investigated the application of texture operators based on spherical harmonics wavelets achieving Local Rotational Invariance (LRI) and Directional Sensitivity (DS) of radiomic features, stressing the importance of the combination of LRI and DS with popular radiomics for classifying 3D textures in particular in medical imaging where local structures of tissues occur at arbitrary rotations [30]. Cohen et al. introduced the Group Equivariant Convolutional Neural Networks (G-CNNs) to transform copies of filters inside CNNs with shared weights to achieve rotational invariance [31]. Winkels et al. used 3D roto-translation group convolutions (G-Convs) in their 3D G-CNNs, with transformed copies of 3D filters to improve performance, sensitivity, and speed of convergence for pulmonary nodule detection in volumetric CT images [32]. Andrearczyk et al. introduced LRI steerable filters and Solid Spherical Energy (SSE) as part of CNNs demonstrating superior performances compared to standard 3D G-CNNs with equivalent number of parameters [33, 34]. All these works were performed by using 3D rotational invariance while trying to keep low the complexity of the resulting DCNNs and avoiding rotational image augmentation pre-processing steps. In regard to the scale invariance, it has to be emphasized that it only exists in discrete multi-step scaling of the original images. Xu et al. used Scale Invariant Convolutional Neural Network (SiCNN) as a DCNN multi-column architecture, with each column focusing on a particular scale, to automatically apply scaling transformation to the CNN layers [35]. Unlike other similar models, the interesting approach is the sharing of filter parameters within the multi-column architecture [35]. In vision research, polar representation has been shown to match human vision in a better way than other image models [36–38]. Esteves et al. presented a Polar Transformer Network (PTN) model to classify objects with rotational invariance, also extending the model to 3D objects with the use of a cylindrical coordinate system [39].

3 Proposed pre-processing and segmentation method

The proposed method consists in a pre-processing step to transform the input volume, followed by a repeated application of the same DCNN with different weights in order to progressively refine the segmentation. Finally, a post-processing step was used to filter excess of segmentation results by using a classical model’s segmentation as well as digital filtering the final segmentation.

3.1 The polar and spherical coordinates system

The method illustrated here features a transformation of the volume from the standard Cartesian spatial coordinate system into the spherical coordinate system. For 2D images, changing the image coordinate system to a polar coordinate system will generate a 2D polar representation, as shown in Fig. 1A. The spherical coordinate system is equivalent to the polar transformation for 3D volumes. An example of the spherical transformation on volumes is shown in Fig. 1B. Most current MRI systems can produce 3D volumes; therefore, the proposed method focused on the transformation of the spherical coordinate system and the application of 3D DCNN models. The coordinate transformation is based on the azimuth angle and polar angle of each point together with its radial distance from an origin, instead of the classical projections of the same point on the x,y,z orthogonal axes. Given an origin O having coordinates (x₀,y₀,z₀) and a point Pₙ having coordinates (xₙ,yₙ,zₙ), the new representation $Tₙ$ of the point Pₙ in the spherical coordinates systems is:

$$Tₙ=f(Pₙ,O) = (rn, θₙ, φₙ)$$

where $rn$ is the distance from $Pₙ$ to $O$, and θₙ and φₙ are respectively the azimuth angle formed on the $x,y$ plane by the projection of $Pₙ$ and $x$ and the elevation/polar angle between $z$ axis and the segment $OPₙ$ [40] (Fig. 3). Values of $rn$, θₙ, and φₙ are uniformly spread on a new volume on the planes $r$, θ, and φ, with θ ranging from $π$ to $−π$, φ ranging from $π/2$ to $−π/2$, and $r$ ranging from zero up the distance from origin to the farthest point in the original volume.

3.2 Baseline network

The CNN chosen as baseline for the proposed method is derived from Myronenko [41], which is based on
Variational Auto Encoder (VAE) U-Net with adjusted input shape and loss function according to the type of tumor to be segmented and the availability of the MRI sequences. Such a model won the BraTS 2018 competition. However, since results do not differ greatly from other simpler networks (e.g., Isensee’s No New Net, second place in BraTS 2018 challenge) [8], and even if the model requires substantial resources, our choice was mainly motivated by the generalization power of Variational Autoencoders. Variational Autoencoders have been shown to be highly reliable; during the training phase, the loss includes a term aimed to let the model learn a regularized Gaussian distribution of the bottleneck layers in the U-Net structure, thus enhancing regularization of the parameters of the latent space and forcing them to be continuous [42]. This kind of regularization improves the generalization of the model over unseen samples instead of matching patterns similar to those already seen during the training, thus performing better than other models on small datasets [43]. Rectified Adam optimizer was used with a learning rate $10^{-4}$ for the training [44]. The implementation of the network was written in Python using Keras on TensorFlow backbone, adapting a model published on Github with minor modification (https://github.com/IAmSuyogJadhav/3d-mri-brain-tumor-segmentation-using-autoencoder-reguarization).

3.3 Cascade CNN models

The pre-processed spherical coordinates volumes are used as input on three CNN models for segmentation, in order to obtain native rotation and scaling image augmentation. An additional CNN Cartesian model is used to filter out false positive results (i.e., redundant segmentation). All of the models share the same CNN layer structure and are used in cascade. They have the same structure of the baseline CNN. Every input volume of the dataset is pre-processed using the output of the previous segmentation (except for the first pass model). The output of each segmentation is used to find the optimal origin points needed to apply the spherical transformation of the next iteration in the cascade CNNs. All the models are trained starting with the weights obtained by the previous model training (except for the first pass model). Since the transformation uses an origin point, after a first segmentation, further steps are needed to project good results backward onto the original volume. In details, the three models are the following: first pass model (origin in the center, random weights initialization), second pass model (four origins taken randomly within the first pass segmentation results, weights taken from the first pass model), third pass (four origins in the centers of the segmented object/objects obtained after the second pass, weights taken from the second pass model). The Cartesian model does not need any origin points and has a random weights’ initialization. These steps refine progressively the results obtained on each iteration. Figure 4 shows an example of the process.

3.3.1 First pass

The first pass model is used to obtain the first weights and to select the next origin point, that should be inside the initial segmentation. In the first approach on the volume segmentation, the origin was chosen around the center of the MRI volume, in order to get an initial segmentation. The first pass segmentation can be also obtained by the Cartesian coordinate systems volume without any transformation, although the use of the transformation on the first pass enables the pre-trained weights to start the second pass model. To obtain a natural image augmentation, four further random origins were added inside the non-zero voxels values of the volume to the first central origin.

3.3.2 Second pass

The second pass segmentation model is trained using four origins falling inside the initial segmented image. To improve the choice of the origin points for the second pass training, the first pass segmentation is filtered from isolated voxels and small objects. The choice of the origin should
preferably be within larger objects, but also in other remote objects not visible enough from the first chosen origin. For such a reason, the choice of the origin points is made according to the following steps:

1. Filter of isolated voxels and small objects, apply “opening” technique (thinning and dilation) to disconnect thin connected objects and remove thin segmentations, fill small holes in the segmentation, identify individual objects from connected voxels.
2. Choose a random origin placed inside the bigger available object segmented. The origin should not be on the border of the object (thinning filter used).
3. Exclude a box having a length of 50 mm for each axis around the origin just considered, and repeat from step 1 until enough numbers of origin points are chosen.
4. In case not enough origin points are selected, repeat from step 1 with less limitation on the small object filtering and thinning procedure.

The value considered for the measure of the small object to be removed is $30 \, \text{mm}^3$ in volume in the first iteration. Successive iterations increase this measure to 100 and 1000 $\text{mm}^3$, beginning to search into the Tumor Core segmentation and then expanding the search to the Whole Tumor segmentation until enough origin points are found.

3.3.3 Third pass

A final selection of the four origin points is performed considering the center of the objects segmented in the second pass. The volume segmentation selects individual objects from connected voxels. If the origin selection is derived from a more sizeable object, greater than $50 \times 50 \times 50 \, \text{mm}$, further origins will be taken as described in the second pass procedure, in order to not exclude remote areas and to reach the four origins needed in this pass.

3.3.4 Cartesian CNN filter model

Since scale and rotation are invariant in the spherical coordinate system models, some useful information within the Cartesian spatial coordinate system can be lost. For such a reason, an additional Cartesian DCNN model has been trained on the same dataset and used to filter out false positive segmentation obtained by the spherical model.

3.4 Image datasets and pre-processing

Two similar resources are used to train and test the proposed method: Medical Segmentation Decathlon (MSD) and BraTS datasets. The Medical Segmentation Decathlon (MSD) Brain Tumors 4D Volumes Dataset [45] was...
used for the training of the models. This dataset includes a total of 750 4D volumes (484 training volumes with labels and 266 testing volumes) consisting of a selection of the BraTS 2016 and 2017 datasets. Every 4D volume contains four MR sequences of skull-stripped and co-registered images of LGGs and HGGs, namely T1, post-contrast T1, T2, and FLAIR. The Training Set contains a manually labelled ground truth. The proposed method uses all the four available sequences, and no further registration was applied since the volumes are already registered, according to the method described by Simpson et al. [45]. The intensity values of the source MR images are not uniform and can vary significantly among different sequences as well as different device setups or different machines. For such a reason, a common approach is to scale the intensities to a normal distribution centered on a value of zero and with a standard deviation of 1, only considering the non-empty voxels. In the proposed method, the normalization was done after the transformation, in order to enhance the target region surrounding the origin point. Furthermore, brain MRI was skull-stripped before performing the transformation and intensity normalization, in order to analyze only the nervous tissue and to assure patients’ anonymization.

3.5 Data augmentation

To augment the initial dataset, a random choice of the origin points during each training iteration was considered to be sufficient. The training phase of the first pass volume considers the origin from the center of volume as well as origins within the whole brain voxels (all the non-empty voxels). In the second and third passes, random origins were taken inside a central region contained in the first segmentation. Thus, the original dataset was augmented dynamically during the training process, having in each epoch a different representation of the original volumes from the training dataset. Even if further augmentation can be done by changing intensity of the voxels, in the proposed method this was already achieved since the normalization of the volume acts like an augmentation process. Indeed, for every different origin choice, a different distribution of the voxels’ histogram is obtained. Thus, during the normalization of the entire volume, intensities of regions vary according to their new extension in the spherical representation, introducing automatically an intensity-based augmentation.

3.6 Label prediction

As mentioned above, the segmentation labels in the used datasets are the three standard labels of the BraTS challenge: (1) Whole Tumor, including all tumor and peritumoral edema, (2) Tumor Core, including necrotic, enhancing, and non-enhancing tumor areas, and (3) Enhancing Tumor area. Label prediction is obtained applying the three models in the same order of the training phase. Firstly, the prediction is done with the first model choosing the origin in the center of the volume. Then, the prediction obtained is used to find the origins for the second model, and then the third model is applied. The second and third models generate local predictions around the chosen origin points of the same volume. The local predictions are then summed up to reconstruct the whole segmentation. The Cartesian CNN model is then used as a filter model, applied on the original non-transformed volume to obtain another segmentation. This last segmentation is used to intersect the spherical segmentation in order to further reduce false positive voxels. Only the Whole Tumor label of the Cartesian model is used as a filter to erase areas wrongly selected as Tumor by the spherical coordinate models, thus meaning that the new WT, TC, and ET labels are obtained by the spherical segmentation excluding all the non-WT area of the Cartesian model segmentation.

3.7 Post-processing

As previously described, the final segmentation considers an ensemble of predictions resulting from different segmentations obtained from the spherical transformed models with all their origin points and projected backwards onto the starting volume. Since the obtained segmentation still contains many false positive voxels, further post-processing steps include the removal of small-sized segmented objects (less than 30 square millimeters in volume size), after removing thin connection between them with an “open” voxelwise operation consisting in thinning and a successive dilation of the objects’ border. The rationale of removing those very small and disconnected regions comes from the observation that human experts–based labels (above all, ET labels) in the whole dataset of ground truth labels are generally exceeding $3 \times 3 \times 3$ voxels. Smaller regions are rarely detected and segmented by experts. Nevertheless, when those small regions appear as false positive in the prediction of a particular patient’s volume, the Dice score for that patient is zero, therefore decreasing the final performance of the method in a considerable way.

4 Experiment and results

4.1 Training phase

The three CNN models were trained in cascade on the MSD Training Dataset. All the 484 MSD Training Set were used for the training phase as follows: 400 input volumes were used for the training and 84 input volumes for the validation. The input volume generator applies the spherical
transformation on every iteration on both the input volume and the ground truth choosing the random origins according to the steps of the method described above for every one of the first, second, and third pass models. So, after the first pass model was completely trained, the obtained segmentation was used to get the origin space for the second pass model. After obtaining the second pass model segmentation, the third pass model was fed with the origins located near the centers of the segmented objects derived from the second pass predictions.

### 4.2 Evaluation

To compare the results to the state-of-the-art on automatic tumor segmentation achieved so far, the proposed method was applied to the BraTS 2019 training and validation datasets and to the online evaluation platform from the CBICA Image Processing Portal (a web platform for imaging analysis, Centre for Biomedical Image Computing and Analytics, University of Pennsylvania, USA: [https://ipp.cbica.upenn.edu](https://ipp.cbica.upenn.edu)) [3, 45, 46]. While the training process uses a Soft Dice Score–based loss (along with the Least Square Errors loss of the VAE reconstruction branch and Kullback–Leibler divergence evaluated on the latent space, as described by Myronenko) [41], the final segmentations were uploaded to the official BraTS 2019 server for evaluation of per class dice, specificity, sensitivity, and Hausdorff distances directly on the BraTS 2019 training and validation datasets. BraTS testing datasets were not used at this stage since the platform only allows to get official evaluation for training and validation datasets. Furthermore, BraTS 2019 testing dataset is not available to those not participating to the competition. The advantage of also evaluating the BraTS 2019 Training dataset is the differentiations of the scores on LGG and HGG groups, while the Validation Set does not distinguish the two glioma groups.

### 4.3 Comparison with baseline

The Cartesian model was used as a baseline for comparison. The spherical model was applied on the same dataset. Then, the post-processing phase was used in the final step to remove false positive voxels. As shown in Table 1, the spherical coordinates model alone did not improve much the overall results obtained with only the baseline model.

On the training dataset, there was a slight overall improvement in the Tumor Core segmentation score (0.011) while WT and ET segmentation scored a similar dice value (difference inferior to 0.008). On validation dataset, a similar trend was observed for TC, while WT results were worse than on the training dataset and ET results improved slightly. Spherical model obtained higher sensitivity on both training and validation dataset compared to the Cartesian models, that means this model includes more true positives compared to the other models. As shown in Fig. 4, segmentation results of the spherical coordinate system were actually selecting the main tumor objects better, but the overall score was worse for the inclusion of many false positive areas selecting normal brain structure having the same characteristics of tumor areas. This is also confirmed by the lower specificity score of the spherical model, as shown in Table 1. These areas were not included in the Whole Tumor areas of the Cartesian model. The final score was then obtained by using the spherical coordinate model filtering the non-WT areas of the segmentation obtained by the Cartesian model and then applying the post-processing to remove small-sized isolated groups of false positive voxels. After this filtering, the final Dice score as well as the specificity improved in all the three classes. The computation of another measure for segmentation accuracy used in the BraTS challenge, the Hausdorff 95, also demonstrated the improvement of the segmentation.

The difference between the score of the two LGG and HGG groups in the BraTS training dataset was also evaluated, showing that the overall TC and ET segmentation were

---

**Table 1** Dice scores, sensitivity, specificity, and Hausdorff 95 of each segmentation label predicted by the models on the BraTS 2019 training and validation datasets: Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET). Bold values highlight the best scores.

| BraTS 2019 Training and Validation datasets | Dice score | Sensitivity | Specificity | Hausdorff 95 |
|--------------------------------------------|------------|-------------|-------------|-------------|
|                                            |            | WT  | TC | ET | WT  | TC | ET | WT  | TC | ET | WT  | TC | ET | WT  | TC | ET |
| BraTS 2019 Training dataset                |            |     |    |    |     |    |    |     |    |    |     |    |    |     |    |    |
| Spherical model                            | 0.882      | 0.803| 0.717 | 0.940 | 0.888 | 0.823 | 0.987 | 0.994 | 0.997 | 7.351 | 7.220 | 5.691 |
| Cartesian (baseline) model                 | 0.890      | 0.792| 0.724 | 0.910 | 0.834 | 0.819 | 0.993 | 0.996 | 0.998 | 7.207 | 8.630 | 5.753 |
| Intersection and post-process              | 0.902      | 0.805| 0.755 | 0.893 | 0.873 | 0.816 | 0.995 | 0.995 | 0.997 | 4.755 | 5.903 | 4.081 |
| BraTS 2019 Validation dataset              |            |     |    |    |     |    |    |     |    |    |     |    |    |     |    |    |
| Spherical model                            | 0.847      | 0.762| 0.704 | 0.941 | 0.836 | 0.792 | 0.985 | 0.994 | 0.997 | 11.550 | 12.220 | 8.496 |
| Cartesian (baseline) model                 | 0.873      | 0.749| 0.690 | 0.915 | 0.764 | 0.749 | 0.992 | 0.997 | 0.998 | 10.250 | 12.790 | 7.925 |
| Intersection and post-process              | 0.879      | 0.764| 0.737 | 0.899 | 0.819 | 0.786 | 0.993 | 0.995 | 0.997 | 8.655 | 10.840 | 6.811 |
better in the HGG dataset compared to the LGG one, while WT was similar (see Table 2).

Table 3 shows the standard deviation and the $t$-test evaluating the improvement of the Dice scores from the spherical models. The spherical transformation model improvement is statistically significant in WT label. The intersection model improvement is statistically significant also in ET label compared to both baseline and spherical model on both datasets, and in TC label on the training dataset compared to the Cartesian baseline.

### 4.4 Application of the model on different resolution volumes

Since the spherical transformation uses natively scaled and rotationally invariant coordinate systems, its application to a different resolution dataset is trivial. However, to validate this, some experiments applying the spherical transformation on randomly rotated and zoomed original dataset were performed, obtaining similar resulting volumes, as shown in Fig. 2. The transformation also introduced different window and level intensity values to the voxels after the normalization step. The BraTS training and validation models had previously been transformed and fed into the new method, obtaining similar segmentation (see Fig. 2). It is worthy to mention that to find the WT label used to filter spherical segmentation, the Cartesian model was applied to the isotropic transformation of the volume, because this model still has the limitation of the isotropic resolution of the dataset.

### 4.5 Hardware and timing

Training was performed on a Fujitsu Celsius R970 workstation (Fujitsu Ltd. Minato, Tokyo, Japan) equipped with 2×24 GB RAM NVidia P6000 GPUs (Nvidia Co., Santa Clara, CA, USA). The complete training of the models required more than 1 month on more than 2000 overall epochs using the 484 MSD Training Set. The segmentation of the full BraTS 2019 Validation Set of 125 patients’ data required 89 min to complete, thus meaning that about 43 s per patient were needed to complete the three-pass segmentation prediction and Cartesian model filtering on the workstation.

### 5 Discussion

Data augmentation techniques have been proven to work well above all in case of extremely limited dataset, but it is usually limited to affine, elastic, and pixel-level augmentation, with some exception in which data augmentation is done generating artificial data mixing volumes from different patients or using GANs [46–48]. Spherical coordinate

### Table 2

|                    | Full training dataset | HGG only | LGG only |
|--------------------|-----------------------|----------|----------|
|                    | WT        | TC        | ET        | WT        | TC        | ET        | WT        | TC        | ET        |
| Spherical model    | 0.882     | 0.803     | 0.717     | 0.885     | 0.854     | 0.804     | 0.868     | 0.627     | 0.417     |
| Cartesian model    | 0.890     | 0.792     | 0.724     | 0.893     | 0.847     | 0.821     | 0.879     | 0.601     | 0.394     |
| Intersection and   | 0.902     | 0.805     | 0.755     | 0.907     | 0.858     | 0.802     | 0.884     | 0.625     | 0.592     |
| post-process       |           |           |           |           |           |           |           |           |           |

### Table 3

| Dice score | St Dev | p-value with Cartesian | p-value with spherical |
|------------|--------|------------------------|------------------------|
|            | WT     | TC         | ET        | WT     | TC         | ET        | WT     | TC         | ET        |
| BraTS 2019 training dataset standard deviation and $t$-test |
| Spherical model | 0.882 | 0.803 | 0.717 | 0.026 | 0.093 | 0.316 | - | - | - |
| Cartesian model | 0.890 | 0.792 | 0.724 | 0.086 | 0.197 | 0.269 | - | - | - |
| Intersection and post-process | 0.902 | 0.805 | 0.755 | <0.001 | 0.024 | 0.019 | <0.001 | 0.212 | 0.001 |
| BraTS 2019 validation dataset standard deviation and $t$-test |
| Spherical model | 0.847 | 0.762 | 0.704 | <0.001 | 0.231 | 0.311 | - | - | - |
| Cartesian model | 0.873 | 0.749 | 0.690 | 0.132 | 0.201 | 0.279 | - | - | - |
| Intersection and post-process | 0.879 | 0.764 | 0.737 | 0.001 | 0.142 | 0.036 | <0.001 | 0.587 | 0.069 |
transformation introduces an extreme augmentation. Each input volume of the brain tumor dataset is modified at every iteration during the training of the model for each origin point randomly chosen. With this augmentation and its invariance to rotation and scaling, the adoption of spherical coordinate system helps to apply the method to datasets coming from environments that use different resolution and settings. It is worth to mention that aligning volumes with a rigid or affine transformation and resampling them to a common resolution are common steps in most image analysis pipelines that aim at removing variability among different institutions settings. However, the approach of resampling volumes down to 1 mm isotropic, as used in BraTS challenges, often results in a loss of large amount of data from the original available volumes. On the other hand, the approach of resampling volumes to higher resolutions without further cropping would need too much memory, becoming unfeasible even on high-performance Deep Learning workstations. Furthermore, even if rotation is not heavily occurring in MRI Scans, and normal brain tissue structures appear very similarly oriented, tumors appear in many different shapes and rotations. Even if the edema included in WT label usually overlaps the anatomic structures surrounding the tumor, TC and ET labels do not appear to be constrained by those structures, and therefore have their own shape and orientation. Spherical

Fig. 5 On the left, an example of a ground truth found in BraTS 2019 dataset, containing a questionable segmentation. The segmentation predicted by the model (on the right) is smoother and appears a more realistic match.
coordinate transformation can preserve the full amount of information included in the original resolution while training the model on new dataset as well as applying it to perform predictions, also taking advantage of rotation invariance, particularly on TC and ET classes. This offers a further confirmation on why the spherical model accuracy is better on those labels instead of the WT label. Indeed, the model learns more intensity and pattern-based features instead of fixed shape features, thus the spherical model has false positive labels that are different from Cartesian models’ ones. The extreme augmentation of the spherical model strengthens the robustness of the DCNN; however, it can occasionally lose information in regard to normal brain structures. Thus, variable accuracies are obtained in both spherical and Cartesian models when applied in distinct regions. The proposed method shows for the first time that the combination of the Cartesian and spherical models together improves the results when compared to the single models alone. Compared to other similar works focused on rotational and scale invariance, the proposed method does not need any change of any Deep Learning frameworks, since it involves only a pre-processing step to transform the coordinate system before using any CNN architecture. Like H-Nets, this method is invariant to any rotation, not only to 90-degree rotations. This method achieves scale invariance, or at least more robustness to scale and resolution variations, like the SiCNN but without the need to replicate columns along CNN layers and copying filters, since the same filters already work on multiscale transformation. The Polar Transformer Network and G-CNNs use a similar approach, but incorporate the transformation into the CNN architectures. PTN also works with 3D objects using a cylindrical coordinate system instead of spherical coordinates. The spherical coordinate systems also give rotational invariance in the elevation angle while cylindrical coordinate system only enables a polar angle invariance on 3D objects. 3D G-CNNs work with 3D objects as well, but do not provide scale invariance.

Even if the spherical transformation produces visually distorted volumes, this distortion is not random but derives from a well-defined function (i.e., the spherical transformation process) and can be restored to the original volumes. By this way, the models are not fed with randomly distorted volumes, rather with the same original volumes represented in a different way. Distortion around origin point is higher compared to the peripheral points; however, center points contain much more information compared to peripheral points; therefore, regions around origin points have better segmentation than peripheral points. For this reason, in the final ensemble, only the region close to origin points was used to compose the final segmentation. This different volumes’ representation allows to use full resolution details around the origin points as well as information from the peripheral areas, while not requiring too much memory for a full resolution volume, nor splitting the volume in patches. Therefore, this technique could be an optimized method of feeding Neural Networks by making better use of the information available from the original volumes.

80/20 split ratio was chosen in the current study as a training/validation split of the original dataset. Different split ratio can be chosen as well as methods, like the k-fold cross validation, to better estimate the performances of the proposed method [49]. However, the training of k-fold cross validation applied on the current method would have been very time consuming; therefore, it will be explored after the parameters’ optimization to improve the training speed.

Our novel method did not score within the best BraTS 2019 competition models. However, more importantly, the proposed method achieved competitive results compared to other DCNN current methods and can be used on different environments and datasets as it is not limited to a specific volume resolution. Furthermore, the neurosurgeon (ADI) involved in the study as well as neuroradiologists who were blinded to the results manually checked the volumes and found that the proposed method corrected some wrong ground truth segmentation on the BraTS’ training dataset (see Fig. 5).

6 Conclusions

A novel approach aimed to change the way DL models are fed is introduced here, with the goal of improving reliability and avoiding limitations of resolution and dataset domain standardization. This new approach includes a natural image augmentation that is a consequence of the spherical transformation technique applied during the pre-processing of input data. This transformation mimics human expert perception, focusing on details and single regions during the segmentation process. Since the spherical coordinate system models are invariant to scaling and rotation, as a drawback they may occasionally lose some useful information of the Cartesian spatial coordinate system. For such a reason, the best performance is achieved when a Cartesian coordinate model is preliminarily used to filter out the non-tumor areas. Results are improved upon the baseline Cartesian model. The proposed model is achieving competitive Dice scores and is not resolution-dependent, therefore can be applied to different datasets without re-training, thus theoretically solving some transfer learning problems related to the domain shifting, at least in terms of image resolution in the datasets. One reason for the limited improvement of the accuracy of the spherical model may be due by the limited number of iterations done so far: no overfitting was observed at the end of a long training phase, so the number of origins used could be still too low compared to the high number of possibilities.
One limit of the methodology is the huge training time that would be required to include a large number of origin points. Further studies will be aimed to optimize the information of the spherical transformation, trying to address this issue and to improve further the performance, by reducing the size of the transformed input and increase the speed and the number of iterations of the training phase. Although only a slight increase in accuracy has been found so far and further studies are required to prove the better performance of the new method to solve the shifting domain problem, experiments based on rescaled and rotated volumes of the BraTS dataset proved the robustness of this method on different settings of the scan parameters of position, resolution, and intensity values; thus, the method is expected to work better and in real-world scenarios. Future studies will also involve collection and manual segmentation of new data using original resolution volumes from the authors’ and other institution, as well as the application of the method to further available datasets.

References

1. Thrall JH, Li X, Li Q et al (2018) Artificial intelligence and machine learning in radiology: opportunities, challenges, pitfalls, and criteria for success. J Am Coll Radiol 15:504–508. https://doi.org/10.1016/j.jacr.2017.12.026

2. Menze BH, Jakab A, Bauer S et al (2015) The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans Med Imaging 34:1993–2024. https://doi.org/10.1109/TMI.2014.2377694

3. Bakas S, Reyes M, Jakab A, et al. (2018) Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge

4. Challen R, Denny J, Pitt M et al (2019) Artificial intelligence, bias and clinical safety. BMJ Qual Saf 28:231–237. https://doi.org/10.1136/bmjqs-2018-008370

5. Allen B, Seltzer SE, Langlotz CP et al (2019) A road map for translational research on artificial intelligence in medical imaging: from the 2018 National Institutes of Health/RSNA/ACR/The Academy Workshop. J Am Coll Radiol 16:1179–1189. https://doi.org/10.1016/j.jacr.2019.04.014

6. Kundel HL, Nomine CF (1983) A visual concept shapes image perception. Radiology 146:363–368. https://doi.org/10.1148/radiology.146.2.6849084

7. Nomine CF, Krupinski EA (1998) Perceptual skill, radiology expertise, and visual test performance with NINA and WALDO. Acad Radiol 5:603–612. https://doi.org/10.1016/S1076-6332(98)80295-X

8. Isensee F, Kickingreder P, Wick W, et al. (2019) No New-Net. pp 234–244

9. Manning D, Ethell S, Donovan T, Crawford T (2006) How do radiologists do it? The influence of experience and training on searching for chest nodules. Radiography 12:134–142. https://doi.org/10.1016/j.radi.2005.02.003

10. Bertram R, Helle L, Kaakinin JK, Svedström E (2013) The effect of expertise on eye movement behaviour in medical image perception. PLoS ONE 8:e66169. https://doi.org/10.1371/journal.pone.0066169

11. Donovan T, Manning DJ (2007) The radiology task: Bayesian theory and perception. Br J Radiol 80:389–391. https://doi.org/10.1259/bjr/98148548

12. Kundel HL, Nomine CF, Conant EF, Weinstein SP (2007) Holistic component of image perception in mammogram interpretation: gaze-tracking study. Radiology 242:396–402. https://doi.org/10.1148/radiol.2422051997

13. Kundel HL, La Follette PS (1972) Visual search patterns and experience with radiological images. Radiology 103:523–528. https://doi.org/10.1148/103.3.523

14. Kundel HL, John Wright D (1969) The influence of prior knowledge on visual search strategies during the viewing of chest radiographs. Radiology 93:315–320. https://doi.org/10.1148/93.2.315

15. Thorpe S, Fize D, Marlot C (1996) Speed of processing in the human visual system. Nature 381:520–522. https://doi.org/10.1038/381520a0

16. Antonopoulos M, Dionysiou D, Stamatakos G, Uzunoglu N (2019) Three-dimensional tumor growth in time-varying chemical fields: a modeling framework and theoretical study. BMC Bioinformatics 20:442. https://doi.org/10.1186/s12859-019-2997-9

17. Torres Hoyos F, Navarro RB, Vergara Villadiego J, Guerrero-Martelo M (2018) Geometrical study of astrocytomas through fractals and scaling analysis. Appl Radiat Isot 141:250–256. https://doi.org/10.1016/j.apradiso.2018.05.020

18. Murray JM, Kaisiss G, Braen R, Klesieck J (2020) A primer on radiomics. Radiolge 60:32–41. https://doi.org/10.1007/s00117-019-06917-w

19. Fan Y, Feng M, Wang R (2019) Application of radiomics in central nervous system diseases: a systematic literature review. Clin Neurol Neurosurg 187:105565. https://doi.org/10.1016/j.clineuro.2019.105565

20. Jang K, Russo C, Di Ieva A (2020) Radiomics in gliomas: clinical implications of computational modeling and fractal-based analysis. Neuroradiology 62:771–790. https://doi.org/10.1007/s00234-020-02403-1

21. Di Ieva A (2016) The fractal geometry of the brain. Springer New York, New York

22. Di Ieva A, Grizzi F, Ceva-Grimaldi G et al (2007) Fractal dimension as a quantitator of the microvasculature of normal and adenomatous pituitary tissue. J Anat 211:673–680. https://doi.org/10.1111/j.1469-7580.2007.08084.x

23. Di Ieva A, Le Reste P-J, Carsin-Nicol B et al (2016) Diagnostic value of fractal analysis for the differentiation of brain tumors using 3-Tesla magnetic resonance susceptibility-weighted imaging. Neurosurgery 79:839–846. https://doi.org/10.1227/NEU.000000000001308

24. Milosevic NT, Di Ieva A, Jelinek H, Rajkovic N (2017) Box-counting method in quantitative analysis of images of the brain. In: 2017 21st International Conference on Control Systems and Computer Science (CSCS). pp 343–349

25. Di Ieva A, Günd S, Grabner G et al (2013) Three-dimensional susceptibility-weighted imaging at 7 T using fractal-based quantitative analysis to grade gliomas. Neuroradiology 55:35–40. https://doi.org/10.1007/s00234-012-1081-1

26. Di Ieva A, Esteban PJ, Grizzi F et al (2015) Fractals in the neurosciences, Part II. Neurosci 21:30–43. https://doi.org/10.1177/1073858413519928

27. Grizzi F, Castello A, Qehajaj D et al (2019) The complexity and fractal geometry of nuclear medicine images. Mol Imaging Biol 21:401–409. https://doi.org/10.1007/s13070-018-1236-5

28. Petrujkić K, Milgošević N, Rajković N et al (2019) Computational quantitative MR image features - a potential useful tool in differentiating glioblastoma from solitary brain metastasis. Eur J Radiol 119:108634. https://doi.org/10.1016/j.ejrad.2019.08.003

29. Worrall DE, Garbin SJ, Turkmukhambetov D, Brostow GJ (2017) Harmonic networks: deep translation and rotation equivariance.
Depeursinge A, Fageot J, Andrarczyk V, et al. (2018) Rotation invariance and directional sensitivity: spherical harmonics versus radiomics features. pp 107–115

30. Cohen TS, Welling M (2016) Group equivariant convolutional networks. Proc Int Conf Mach Learn

31. Winkels M, Cohen TS (2019) Pulmonary nodule detection in CT scans with equivariant CNNs. Med Image Anal 55:15–26. https://doi.org/10.1016/j.media.2019.03.010

32. Andrarczyk V, Oreiller V, Fageot J, et al. (2019) Solid spherical energy (SSE) CNNs for efficient 3D medical image analysis. Irish Mach Vis Image Process Conf

33. Andrarczyk V, Fageot J, Oreiller V, et al. (2019) Exploring local rotation invariance in 3D CNNs with steerable filters. Proc Mach Learn Res

34. Xu Y, Xiao T, Zhang J, et al. (2014) Scale-invariant convolutional neural networks. ArXiv

35. Edelman JA, Mieses AM, Konnova K, Shiu D (2017) The effect of object-centered instructions in Cartesian and polar coordinates on saccade vector. J Vis 17:2. https://doi.org/10.1167/17.3.2

36. LIN C-S, CHEN H-T, LIN C-H, et al (2005) Polar coordinate mapping method for an improved infrared eye-tracking system. Biomed Eng Appl Basis Commun 17:141–146. https://doi.org/10.4103/S1016237205000226

37. Cantoni V (1994) Human and machine vision : analogies and divergencies. Springer

38. Cantoni V (1994) Human and machine vision : analogies and divergencies. Springer

39. Esteves C, Allen-Blanchette C, Zhou X, Daniilidis K (2018) Polar transformer networks. ArXiv

40. Eric W. Weissstein (2005) Spherical coordinates -- from MathWorld-a Wolfram web resource

41. Myronenko A (2019) 3D MRI brain tumor segmentation using autoencoder regularization. In: Brainlesions: glioma, multiple sclerosis, stroke and traumatic brain injuries. BrainLes 2018. Lecture Notes in Computer Science. Springer. Cham, pp 311–320

42. Kingma DP, Welling M (2014) Auto-encoding variational Bayes. ICLR 2014 - Conf Track Proc

43. Doersch C (2016) Tutorial on variational autoencoders. ArXiv

44. Liu L, Jiang H, He P, et al. (2019) On the variance of the adaptive learning rate and beyond. ICLR 2019

45. Simpson AL, Antonelli M, Bakas S, et al. (2019) A large annotated medical image dataset for the development and evaluation of segmentation algorithms. ArXiv

46. Nalepa J, Cwick M, Dudzik W, et al. (2019) Data augmentation via image registration. In: 2019 IEEE International Conference on Image Processing (ICIP). IEEE, pp 4250–4254

47. Nalepa J, Marcinkiewicz M, Kawulok M (2019) Data augmentation for brain-tumor segmentation: a review. Front Comput Neurosci 13:. https://doi.org/10.3389/fncom.2019.00083

48. Ribalta Lorenzo P, Nalepa J, Bobek-Billevich B et al (2019) Segmenting brain tumors from FLAIR MRI using fully convolutional neural networks. Comput Methods Programs Biomed 176:135–148. https://doi.org/10.1016/j.cmpb.2019.05.006

49. Xu Y, Goodacre R (2018) On splitting training and validation set: a comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. J Anal Test 2:249–262. https://doi.org/10.1007/s41664-018-0068-2

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