Custom administering attention module for segmentation of magnetic resonance imaging of the brain

Nagaveni B. Sangolgi¹, Sasikala²
¹Department Electronics and Communication, Faculty of Engineering and Technology, Sharnbasva University, Kalaburagi, India
²Godutai Engineering College, Osmania University, kalaburagi, India

Article Info

Article history:
Received Jul 27, 2022
Revised Feb 24, 2023
Accepted Mar 24, 2023

Keywords:
Custom administering attention module
Feature maps
Long-range dependency
Magnetic resonance imaging
U-Net model

ABSTRACT

Taking into account how brain tumors and gliomas are notorious forms of cancer, the medical field has found several methods to diagnose these diseases, with many algorithms that can segment out the cancer cells in the magnetic resonance imaging (MRI) scans of the brain. This paper has proposed a similar segmenting algorithm called a custom administering attention module. This solution uses a custom U-Net model along with a custom administering attention module that uses an attention mechanism to classify and segment the glioma cells using long-range dependency of the feature maps. The customizations lead to a reduction in code complexity and memory cost. The final model has been tested on the BraTS 2019 dataset and has been compared with other state-of-the-art methods for displaying how much better the proposed model has performed in the category of enhancing, non-enhancing and peritumoral gliomas.

This is an open access article under the CC BY-SA license.

1. INTRODUCTION

Gliomas are a type of tumor that is an outgrowth of neural cells in the spinal cord and the brain. In the beginning stages, the glial cells, that is, the starting glioma cells surround the nerve and help in functioning. This is the reason why they are so hard to early diagnose this type of cancer. A physical or medical exam is required to check for any coordination deprivation, reflex and speech mismanagement, and other behavioral patterns, such as a magnetic resonance imaging (MRI), computed tomography (CT) scan, or biopsy. The problems faced by the state-of-the-art algorithms for MRI image segmentation are the background noise, lesions being too pixelated, and the gliomas being too fused into the brain cells that segmentation is not possible. Many algorithms have found solutions to these problems; most of them employed convolutional neural networks and variations of the same. The research is U-Net encoder-decoder dense predictor, which is a fully convolutional neural network that employs end-to-end semantic segmentation mechanism [1]-[4]. Other similar methods are [5]-[12] which uses the 3-D U-Net architecture and encoder-decoder, with [13] using skip connections. Another problem pops out of the methods that are the problem of small-sized kernels caused due to the large distance connection between voxels. This leads to suppressing of under or oversegmentation. One way to solve this is to stack up the convolutional layers but computational and memory costs increase exponentially. Yu and Koltun [14] have found a way to solve this by implementing dilated convolution that helps in expanding the receptive of the local feature scale. The non-local neural network implementation of double matrix multiplication that provides the pixel relation in feature and other maps [15].

Journal homepage: http://ijres.iaescore.com
With that, the weight coefficient is found that helps in obtaining global information. They developed a 3-D U-Net combination with a self-attention method that provides global information by making the attention-grating module that is a part of embedded skip connections [16]. Even with so many methods, the problems of matrix multiplication framework, that is, instability of the attention map and the increased code complexity. The code complexity reaches out to be $O(n^2)$ for 2-D images with a huge GPU memory consumption. Moreso, the 3-D images have it at higher stakes as a context prior layer is referencing the attention mechanism.

Now the proposed solution of this paper aims to solve the problems posed. Hence, the model named custom administering attention module, which uses a convolution that forms an attention map using the MRI scans as input, followed by the serial convolution along with truth accuracy guiding administering. Figure 1 shows the visual representation of the proposed model. Then, an endo-class updating method is introduced that makes the feature map reworked and self-categorizes without undergoing a second matrix multiplication, which helps in reducing code complexity and computational costs.

![Diagram of the proposed model](image)

**Figure 1.** This image is a visual representation of the proposed model

In explaining the Figure 1, the purple arrows depict matrix convolution with the leftmost block being the dropout and the red block being the soft-max. This paper is divided into 5 sections. Section 1 handles the introduction; in section 2 handles all the related work and the beginning of the proposed model. Section 3 is the mathematical analysis of the mode while section 4 displays the results and comparisons with other state-of-the-art methods. In section 5 is the conclusion and future scope, which is then followed by the reference section.

2. RELATED WORK

The MRI scans obtained in hospitals are filled with noise and diffused pixels that trained professionals can only demarcate. For algorithms to do the same, methods such as clustering [17], thresholding [15], and so on. Ronneberger et al. [3] handle 2-D and 3-D images using U-Net-based skip connections. The features are extracted and classified based on pixels. The researchers [18]–[20] recover these features by modifying the basic unit and then [21], [22] extract the richer features.

Here this survey checks on the different large-scale training samples that have high labeling medical costs and little supervision which is an issue in segmenting and classifying images [23]. There is another survey on the different types of gliomas and how different machine-learning models are there that have effectively segmented the scans of these lesions [24]. They developed another peritumoral paper that uses an algorithm that segments those tumors with a specified radius and has also been tested on 285 mp MRI scans [25]. This paper gives a survey that samples different deep-learning methods for brain lesion scan classification as well as compares different state-of-the-art methods [26]. They use a multimodal disentangled variational...
autoencoder (MMD-VAE) that extracts features from preoperative multimodal MRI scans [27]. It quantizes the radionics and then the latent representations of the variational autoencoder are disassembled into distinctive and common representations to obtain the shared and complementary data among modalities. To increase effectiveness, it uses cross-modality reconstruction and common distinctive loss. Then shapey additive explanations (SHAP) are employed to implement quantitatively interpret and analyze the important features to grade them. Then the method is tested on two public databases and provided good accuracy results. CCNet is great in reducing complexity that decreases the attention map’s two dimensions that include 3-D images and local map feature map [28] is chosen in such a way that makes where $M \in R$ and follows $X \times Y \times Z$, where it means spatial height, spatial width, and channel number respectively. Then, the attention model applies the $1 \times 1$ convolutions, that is, $Y_p$, $Y_q$ and $Y_o$ respectively and a reshape operation that will provide the matrices \{P, Q, O\} and it belongs to $R^{Z' \times X \times Y}$, where $Z'$ is the channel number after the convolution and the three matrices can be written as (1):

$$P = R\left(Y_p(M)\right), \quad Q = R\left(Y_q(M)\right), \quad O = R(Y_o(M))$$

then, the SoftMax normalization function $f$ is used and matrix multiplication to obtain the attention map $A \in R^{N \times N},$

$$A = f(P \times H^T)$$

then, using reformation and multipication function, a similarity matrix $B$ will be constructed and the result comes out to be.

$$F_R = R(B \times A) + X$$

There are some drawbacks to this formulation, one being that if the input is larger than expected the complexity skyrockets and the attention map’s memory cost is heavy. Reduction of the attention map size is very important and thus increases effectiveness and reduces code complexity. Jiang et al. [29] is useful for solving these new issues as it applies context prior layer and uses convolution to remake the map $Q$ that has the same size as $A$.

$$Q = R(\text{conv}(M))$$

ground truth $G_T$ is now encoded with the one-hot encoding and the formula obtained is $D \in R^{X \times Y \times CI}$, $CI$ denoting the number of classes. Then affinity map is obtained by (5).

$$S \in R^{CI \times X \times Y}$$

$$D = \text{onehot}(G_T)$$

$$S = D \times D^T$$

Context prior layer is used for supervising the affinity map and this increases accuracy and reduces instability. To obtain the leftover information, the size is reduced using the CP map while it groups up similar pixels together, but this increases computational complexity and cost. So now the CP turns the feature map $M$ to $Q$ using a convolution layer and this changes the physical layers and the problems that arise due to this conversion make the first dimension of the CP map the channel dimension. This is not the same as the relationship between two points’ spatial dimensions. The only problem here is that the variables are shared before the update has even taken place. All attention encoder-decoder map, an attention U-Net that employs an attention gate module that looks out for specific shapes and sizes and self and multi-scale attention maps [29]–[32]. With the help of attention weight, the supervision information is used to reduce the code complexity of matrix multiplication. To maintain consistency of a section, endo-class distance is needed, and exo-class optimization helps in amplifying the differences between the classes. Ming et al. [33] show a triple loss based on endo and exo classes. The endo-class consistency between pixels as well as the reworked exo-class distance is managed using semantic segmentation [34] and has been of great help in solving this dilemma of inconsistency and obtaining boundary knowledge. The other problems are solved using global semantic information and then obtain the prototype vector for each type of class, followed by the re-weighting and auto-correlation of the attention mechanism is helped in optimization.
3. MATHEMATICAL ANALYSIS

3.1. Overview

The proposed model has used the category-guided attention U-Net as a base reference to formulate the solution. Input is sent through an attention convolution path to generate an attention map and this map is used for helping in updating the endo-class process and exo-class distance optimization. In the entire process, 3 losses occur, namely main loss (\(L_M\)), attention map loss (\(L_A\)), and inter-class loss (\(L_I\)), the main being corrected using SoftMax dice loss and the adoption of mean squared error (MSE) loss for attention map supervision. Then there is reconstruction, reformation, and involvement of unconventional attention convolution path for the avoidance of any problems and issues that may result from the after-mentioned optimizations. There are two stages of optimizing loss. The first 20 passes involve the addition of \(L_A\) and \(L_M\) while the next 20 passes calculate the total loss sum. A weight is added to keep the value of \(L_I\) at par with the others, while the weights added to the other losses are constant and the same. The feature map is updated based on categories which further helps in simplifying the code.

3.2. Updating the endo-class

The endo-class updating method is a feature of the proposed model that works on the feature map, having total \(X \times Y\) vectors vectored pixels of \(1 \times Z\), as \(M \in R^{X \times Y \times Z}\), the veil \(V_i \in R^{X \times Y}\) that is referenced to all parts of the attention map when it's dismantled. \(Q_i\) is a prototype vector for veiled average pooling which is shown as (6):

\[
Q_i = \frac{1}{\sum_{j=1}^{X \times Y}} \sum_{i=1}^{X \times Y} v_i \times w_i
\]

(6)

here, \(w_i\) is the weight in veil \(i\) and \(v_i\) is the feature map’s vector. Then the prototype vector is mapped back to the size of the lower feature map according to the veil as (7):

\[
m_i = \text{map}(Q_i, V_i)
\]

(7)

here, \(m_i\) is the mapped feature map vector. Finally, all section maps are added to aggregate pixels of the same type.

3.3. Updating the exo-class

To categorize and segregate the same pixels, this method has been introduced which optimized the space between these pixels. The Euclidean distance between the pixels is calculated and then all of them are summed up to represent the total exo-class feature map distance, which is shown as (8):

\[
L = \sum_{0<i<j<C} E(Q_i, Q_j)
\]

(8)

here, \(E\) is the Euclidean distance between vectors. To increase this value, an exo-class loss is also introduced as (9):

\[
l_{exo} = \log(\frac{1}{1+l})
\]

(9)

the less \(l\) can be further reduced in the training for distance optimization.

3.4. Attention map update

This will automatically result in the attention map by using the given input, using serial convolutions use Figure 2. The size of the map is reduced to \(X \times Y \times C\) from \(N \times N\), where \(C\) is the number of classes. The one-hot form is formulated by encoding the down-sampled ground truth and then section-guided map \(S_{G_M}\) and is the same size as \(A\). The \(1 \times C\) vectors are valued high (at 1) while the rest are valued low (at 0) ad each stream of \(S_{G_M}\) is part of a category of pixel distributions. Then the section-guided map trains the attention map. The attention map representation is given by (10).

\[
A = \text{conv}(J)
\]

(10)

4. RESULTS AND EVALUATION

For testing purposes, the model has been tested using the BraTS 2019 [3] dataset. The results were then compared with the result of the paper [35], that is, the U-Net++ model, [36] the U-Net architecture model,
and [3], [37–39] the one inspired by U-Net. The BraTS 2019 [3] dataset evaluates the state-of-the-art methods for image segmentation of the MRI of the brain. It is a multi-institutional pre-operative scan and focuses on the segmentation of intrinsically heterogeneous gliomas. Table 1 given below compares the value sensitivity and specificity, dice score, and 95% Hausdorff distance for the enhancing, non-enhancing and peri-tumoral gliomas (acronyms for them being ETA, NTA, and PTA respectively). The best values attained for each category for different parameters are marked in bold. Table 1 compares the values of sensitivity and specificity of the various state-of-the-art methods, the reference paper used, and then the proposed model of this paper. Figure 2 shows a sensitivity comparison of the various state-of-the-art methods along with the base paper and or proposed model. Figure 3 shows a specificity comparison of the various state-of-the-art methods, the base paper, and or proposed model. The proposed model of this paper has performed the best in almost all the categories and parameters.

Table 1. This table compares the values of sensitivity and specificity of the various state-of-the-art methods, the reference paper used, and then the proposed model of this paper.

<table>
<thead>
<tr>
<th>Method</th>
<th>ETA</th>
<th>NTA</th>
<th>PTA</th>
<th>ETA</th>
<th>NTA</th>
<th>PTA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amian and Soltaninejad [37]</td>
<td>0.68</td>
<td>0.82</td>
<td>0.74</td>
<td>1</td>
<td>0.99</td>
<td>1</td>
</tr>
<tr>
<td>Wang et al. [38]</td>
<td>0.766</td>
<td>0.897</td>
<td>0.826</td>
<td>0.998</td>
<td>0.995</td>
<td>0.996</td>
</tr>
<tr>
<td>Hamghalam et al. [39]</td>
<td>0.769</td>
<td>0.913</td>
<td>0.777</td>
<td>0.999</td>
<td>0.994</td>
<td>0.998</td>
</tr>
<tr>
<td>Micallef et al. [35]</td>
<td>0.723</td>
<td>0.867</td>
<td>0.763</td>
<td>0.998</td>
<td>0.994</td>
<td>0.997</td>
</tr>
<tr>
<td>Proposed solution</td>
<td>0.922</td>
<td>0.88</td>
<td>0.888</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
</tr>
</tbody>
</table>

Figure 2. Sensitivity comparison of various state-of-the-art methods, the base paper, and or proposed model

Figure 3. Specificity comparison of various state-of-the-art methods, the base paper, and or proposed model
Hamghalam et al. [39] has proven that glioma pixels are more than the half of the population in the pathological sample of pixels, making the NTA have the maximum specificity value. ETA has performed poorly due to the low-grade glioma cells not being so clearly demarcated from the brain cells that segmentation is nearly impossible. High-grade gliomas on the other hand have great contour shapes and are well defined to give an output of higher values. The dice coefficient values show that the segmentation accuracy for the proposed model is quite high. Hausdorff distance for NTA and PTA has high amounts of deviations along with great outliers; hence, we take 95% of the largest segmentation error. Table 2 has 10 images from the BraTS 2019 dataset that the proposed solution has worked on. As can be seen, the result of the model and the ground truth are quite close, making it highly accurate.

| Table 2. 10 images from the BraTS 2019 dataset that the proposed solution has worked on |
|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|
| The images from the BraTS 2019 dataset (MRI scans) | The ground truth of the glioma areas | The results of the base reference paper | The final results were obtained by using the images of row ‘a’ as input |
| ![Image 1](image1.png) | ![Image 2](image2.png) | ![Image 3](image3.png) | ![Image 4](image4.png) |
| ![Image 5](image5.png) | ![Image 6](image6.png) | ![Image 7](image7.png) | ![Image 8](image8.png) |
| ![Image 9](image9.png) | ![Image 10](image10.png) | ![Image 11](image11.png) | ![Image 12](image12.png) |
| ![Image 13](image13.png) | ![Image 14](image14.png) | ![Image 15](image15.png) | ![Image 16](image16.png) |

5. CONCLUSION AND FUTURE SCOPE

The existing system used a two-way pipeline with a full-resolution, input image. Hausdorff distance values for this model have not been the best, especially for non-enhancing tumors, due to its diffused pixels. It has performed much better in those areas where the proposed solution has not done its best. As it can be concluded, the custom administering attention module has outperformed several state-of-the-art algorithms in terms of sensitivity, specificity, dice coefficient, accuracy, and efficiency. The BraTS 2019 database is a great choice for training due to its extensive samples and the proposed model has provided almost accurate results in comparison with the ground truth. There are some ways to improve this model. One of them is to involve pipelining and clustering pixels and making it more intense, thus helping in segmenting the more diffused NTA and peritumoral cells.

Custom administering attention module for segmentation of magnetic ... (Nagaveni B. Sangolgi)
REFERENCES


BIOGRAPHIES OF AUTHOR

Nagaveni B. Sangolgi  Working as assistant professor in the Department of Electronics and Communications Engineering at Sharnbasva University, Kalaburagi. She has received her M-Tech. She has published eight International Conferences. Her area of interest is image processing. She can be contacted at this email: nagvenimk289@gmail.com.

Dr. Sasikala  completed bachelor of electrical in electrical engineering from Osmania University Hyderabad in the year 1985 with first class distinction. Obtained ME in the year 1987, in power systems from the Osmania University and awarded gold medal for standing first in the University. Completed Ph.D. from JNTU Hyderabad in electrical engineering in the year 2008. Working as Professor/Principal in various colleges from 2008 and currently working as principal of Godutai engineering College for Women, Sharnbasva University. Kalaburagi’ since 2014. Has 3 decades (30 years) of teaching experience. Published about 50 papers in various national and international journals and presented about 26 papers in both national and international conferences. She can be contacted at this email: sasi_mun@rediffmail.com.