International Journal of Scientific & Engineering Research, Volume 11, Issue 5, May-2020
ISSN 2229-5518

Guiding Mechanism for encoder-decoder leverage for segmentation of stroke lesions

Fredy Malack Sabuni[1], Prof. Li Fenglian[1*], Ntikurako Guy Fernand Kazihise[1]

1College of Information and Computer Science, Taiyuan University of Technology, Taiyuan, Shanxi, China
Corresponding author: Li Fenglian,ghllfl@163.com

Abstract— Semantic segmentation for medical images is a crucial trait in diagnosing, treating, and tracking various diseases. Traditional interpretations are usually applied in clinical practice, but they tend to consume time during the procedure and vulnerable to inter-intra observer variation. Automation segmentation is a method that can quickly find the lesion area by using image signal processing technology. It has become a research hot point in lesion segmentation of medical images in recent years. The latest advances in CNN have added rise to techniques with excellent scores, which are found in most proposed frameworks for biomedical segmentation. This phenomenon can also be seen from the new methods in MRI for the segmentation of stroke lesions. More specifically, U-shaped U-Net-based architectures are thriving for semantic segmentation. This paper is inspired by the guiding mechanism, which has been successful in sequence to sequence challenges such as visual image caption generation to segment stroke lesion. The guiding mechanism is a technique in deep learning, whereby the model/network is being guided to focus on specific features over others during data processing. In this paper, model performance assessment is tested against the ATLAS dataset that consists of patients with stroke lesions. Dice score coefficient (DSC) is used to compare the proposed method's model output against other benchmarks previously proposed for the segmentation of stroke lesions, which has proven to have a better performance compared to previous studies.

Index Terms— Stroke lesions segmentation, U-Net, Convolutional Neural Networks, guiding mechanism.

1 Introduction

It is without question that stroke is one of the life-threatening diseases, which causes a lot of deaths yearly [1]. This is also a significant cause of long-term impairment, leading to impaired muscle function, sensory or emotional disturbances, language comprehension problems, and memory loss. Acute ischemic strokes, caused by an artery obstruction, account for 87 per cent of all strokes. The area in the brain that has been compromised by blood clotting is usually denoted as penumbra (tissue at-risk) [2], and it is the area of focus during the process of curing strokes. There are various types of brain imaging that have been used to assess patients with strokes, but the most common ones are Magnetic Resonance Imaging (MRI) and Computerized Tomography (CT) [3]. Recently, MRI has been the most preferred approach compared to CT because MRI [4][5] represents different successions such as T1 & T2 weights, FLAIR, DWI images that are appropriate for the position and distribution of ischemic stroke lesions. For some time now, numerous techniques for automatic segmentation of stroke lesion were built based on unsupervised clustering of k-means [6], active learning approaches [7] or more advanced [8], and recently, convolutional neural networks (CNN) demonstrates the high performance of the function and outperforms standard segmentation approaches on benchmark datasets [9] [10]. CNNs have been studied in biomedical imaging with many uses in the area of neuroimaging. While a lot of medical images (MRI or CT) are still in 3D dimensions, the majority of CNN work has addressed the issue by using slice-by-slice 2D image analysis. This paper, through spatial and channel-wise attention, allows learning to concentrate without external interference on goal structures quickly. And an implemented technique that involves the integration of spatial & channel-wise attention module is elaborated. The attention module addresses the general CNN encoder-decoder problem in some degree. Nevertheless, the proposed method does not present substantial operating computation cost and does not include an enormous number of parameters, as in the framework's multi-model. In this task of stroke lesion segmentation on MRI, validating the process, the ATLAS dataset is used. It is available to the public. The result of the proposed method improved the efficiency of an ordinary CNN encoder-decoder architecture by successfully modeling rich context-dependence over local characteristics.

2 Related works

2.1 Stroke lesions Segmentation

Recently, there has been a close follow-up in lesion segmentation, whereby multiple methods for stroke segmentation have been presented. For instance, using a Markov Random Field (MRF) models [11], Extra tree forest [6], unsupervised k-means clustering [6]. In the research proposed by Maier etc. [12], a stroke segmentation method based on local features extracted from multimodal (MRI), and a support vector machine (SVM) classifier was further trained to segment the lesions. The random-forest-based method was also used to identity the sub-acute ischemic stroke lesions, which was among the top-ranking methods in Ischemic Stroke Lesions Segmentation (ISLES) challenge in 2015 [13]. Although the above-proposed methods have been successful, it appears that there are still some limitations in the capabilities of segmenting stroke lesions. However, convolutional neural networks (CNNs) have achieved state-of-art performance in a field of visual recognition tasks, becoming very popular due to

International Journal of Scientific & Engineering Research, Volume 11, Issue 5, May-2020
ISSN 2229-5518

their powerful, nonlinear feature extraction capabilities. This has influenced the advancement in segmentation [12]. Despite of CNN strong representation power, these multi-scale approaches lead to excessive use of information flow. Furthermore, CNN encoder-decoder architecture for some tasks such as medical image segmentation, which are pixel-wise segmentation, is proved to be insufficient.

2.2 Attention Mechanism

Attention mechanism in deep learning has improved the success of various tasks in recent years and continues to be an omnipresent component in state-of-art models. Attention mechanism targets at emphasizing local regions captured in local features, and filtering irrelevant information transferred by global features, and improving the modeling of long-range dependencies. Integration of attention modules has been used in Image QA (Question-Answer) as proposed by the author [14], where he found the concatenation of image features and bag of words features worked the best. As for VQA (Video Question Answer), Zhu et al. [15] adopted the "soft" to merge the image features region. Although the attention mechanism is becoming popular on many vision problems, the literature on medical image segmentation with attention remains deficient, with a simple attention model [16][17]. For instance, the attention module has been applied by Wang et al. [13] for prostate segmentation on Ultrasound images. Most lately, the attention gate module [18] has been integrated into the skip connection of UNET to complement information from the encoder.

3 Methodology

This paper investigates automatic stroke lesion segmentation by combining deep convolutional neural networks with guiding mechanisms, which are a technique in deep learning whereby the model/network is being guided to focus on specific features over others during data processing. We employ convolution-based U-Net to extract global context, then fuse features obtained from convolutional layers into attention modules that contain spatial and channel self-attention modules.

3.1 Attention methods

Attention method, as it has been introduced in the different research study, is a technique which focuses on filtering unnecessary features from visual input data [19-22]. Attention methodology is beneficial, especially when dealing with complex visuals with a lot of features. It can be used to eliminate features that are not important. It is benefit for improving and simplifying the execution process. Just like how convolutional networks have developed the whole process of computer vision, one can also view attention technique as one of the means which positively affects the field as well. Among other areas, the attention mechanism has demonstrated a better performance in dealing with medical visuals such as medical images analysis segmentation tasks [20]. As discussed previously, due to the deficiency of standard convolution models, the capability of modeling feature representations is limited. To address the challenge, this paper introduces the position and channel attention as it has been proposed by Shyr A et al. [20]. Whereby the position & channel attention module is responsible for integrating global dependencies to their local features, and together, they form the idea of guiding mechanism, which is the cornerstone for this paper.

3.2 Position attention method

The position attention module possesses the merit for assisting on encoding a more extensive range of contextual information into local features. It is benefit for enhancing the representation capacity. Let's denote a local input features as $\boldsymbol{I} \in \mathbb{R}^{C\,X\,H\,X\,W}$ where $\boldsymbol{C}, \boldsymbol{W}, \boldsymbol{H}$ represent the channel, width and height dimensions respectively. The upper branch $\boldsymbol{I}$ is passed through convolutional block results into feature maps $\boldsymbol{K}$ and $\boldsymbol{P}$, respectively, where$\{\boldsymbol{K}, \boldsymbol{P}\} \in \mathbb{R}^{C\,X\,H\,X\,W}$.Then reshape them to $\mathbb{R}^{C\,X\,V}$, where $\boldsymbol{V} = \boldsymbol{H}\,\boldsymbol{X}\,\boldsymbol{W}$ is the number of pixels. Then a matrix multiplication is applied between the transpose of $\boldsymbol{K}$ and $\boldsymbol{P}$, and the SoftMax layer is used to compute spatial attention map $\boldsymbol{S} \in \mathbb{R}^{V\,X\,V}$:

$$\boldsymbol{s_{ji}} = \frac{\exp(\boldsymbol{K_i} \,.\, \boldsymbol{P_j})}{\sum_{i=1}^{N} \exp(\boldsymbol{K_i} \,.\, \boldsymbol{P_j})} \tag{1}$$

Where $\boldsymbol{s_{ji}}$ measures the $\boldsymbol{i^{th}}$ position's impact on $\boldsymbol{j^{th}}$ position. The input $\boldsymbol{I}$ is fed into a different convolutional block to generate a new feature map $\boldsymbol{D} \in \mathbb{R}^{C\,X\,H\,X\,W}$and reshape to $\boldsymbol{S} \in \mathbb{R}^{V\,X\,V}$. Then it is multiplied by a scale parameter $\boldsymbol{\beta}$ and perform an element-wise sum operation with $\boldsymbol{I}$ to obtain an output $\boldsymbol{F} \in \mathbb{R}^{C\,X\,V}$ as shown below:

$$\boldsymbol{F_j} = \beta \sum_{i=1}^{N} (\boldsymbol{s_{ji}} \boldsymbol{D_i}) + \boldsymbol{I_j} \tag{2}$$

As in [20], the value of β is initialized as 0, and it is gradually learned to give more value to spatial attention map.

3.3 Channel attention method

In channel attention module, it is mainly built for providing interdependence between channels. The input $\boldsymbol{I} \in \mathbb{R}^{C\,X\,H\,X\,W}$ channel attention map is calculated as. $\boldsymbol{X} \in \mathbb{R}^{C\,X\,C}$ Specifically, the original shape $\boldsymbol{I}$ is changed to $\mathbb{R}^{C\,X\,V}$, and then a matrix multiplication between $\boldsymbol{I}$ and the transpose of $\boldsymbol{I}$ is performed to obtain a channel attention map: $\boldsymbol{X} \in \mathbb{R}^{C\,X\,C}$

$$X_{ji} = \frac{\exp(\boldsymbol{I_i} \,.\, \boldsymbol{I_j})}{\sum_{i=1}^{N} \exp(\boldsymbol{I_i} \,.\, \boldsymbol{I_j})} \tag{3}$$

Where $\boldsymbol{X_{ji}}$ measures the $\boldsymbol{i^{th}}$ channel impact on $\boldsymbol{j^{th}}$ the channel.

International Journal of Scientific & Engineering Research, Volume 11, Issue 5, May-2020
ISSN 2229-5518

This is then multiplied by a transposed version of $\boldsymbol{X}$ and $\boldsymbol{I}$, whose result is reshaped to $\mathbb{R}^{\boldsymbol{C\,X\,H\,X\,W}}$. Then the final channel attention map is obtained as:

$$E_j = \beta \sum_{i=1}^{N}(X_{ji}I_i) + I_j \qquad (4)$$

As in [20], the value of β gradually learns a weight from 0. The channel attention module then selectively aggregates features of all the channels into the original ones, highlighting class-dependent feature maps and improving class-dependent feature discriminability.

3.4 Proposed architecture

During data processing, Guiding Mechanism (Partial Attention & Channel Attention) is being used as aggregators because they're able to take full advantage of a vast range of contextual information. The outputs produced by Partial Attention & Channel Attention are fused into convolution layers using an element-wise sum method. Our proposed architecture as illustrated in Figure 1. First, an input image is being passed through convolutional layers, batch normalization for improving the performance of an ANN, then ReLu which introduces the positive part in the network. Features extracted from convolutional blocks are also passed through the Guiding block, which assists in amplifying vital features while diminishing unnecessary ones. Finally, the details obtained from convolutional layers and the ones generated through the guiding block are all being concatenated together. This whole introduced method helps to minimize the number of parameters being extracted but improve realization accuracy.

4 Experiment Setup

4.1 Dataset and preprocessing

In this study, the experiment is performed on the ATLAS dataset. ATLAS (http://fcon_1000.projects.nitrc.org) is a public database consisting of 304 T1w anatomical MRIs of individuals with chronic stroke collected by research groups worldwide from the ENIGMA Stroke Recovery Working Group consortium. To account for potential confounding factors, only MRIs with 1 mm isotropic voxels are included, and all MRIs were collected from 3T scanners. Only one MRI per individual (no inclusion of longitudinal data) is included. 181 T1w anatomical MRIs (100 left hemisphere stroke (LHS), 81 right hemisphere stroke (RHS)) from a total of eight different scanners are included in the current analyses. The dataset of ATLAS contains several stroke visuals of different patients, and the visuals vary in depth based on the slice used to reconstruct an image. During the pre-processing step, it is found that the 3D raw size of the image, which is 256 * 256 * 120, is too huge to be processed because it requires a lot of computational power. So, it is necessary to evaluate the image size for efficiently training the model without losing original data. By experiments, it is found that an image 3D size of 128 *128 * 120 can preserve its original data.

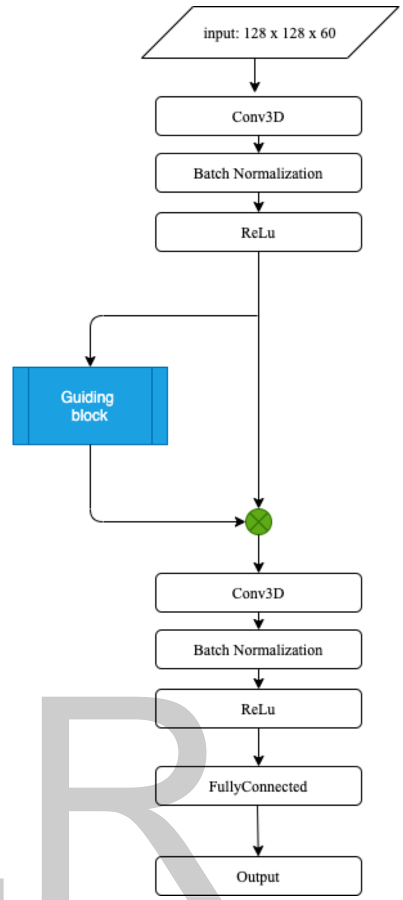

Figure 1: Proposed architecture

4.2 The Evaluation

To evaluate the performance, this study selects 85% of the dataset for training and 15% for validation. The model is trained 5 times with random initialization. In the investigation, it was concluded that the increase in the number of epochs improves the performance of the model. This study used 10, 30, 50, 100, and 200 as training epochs during the experiments. For the baseline approach, this study uses the proposed methodology without the introduction of attention mechanism. Then the effect of the attention method Is evaluated when introducing the original 3D U-Net architecture and observing the performance change. DSC can express the relationship between the actual area and the predicted one. Dice coefficient (DSC) [28] is used in this paper to calculate the performance of the network quantitatively. DSC can be calculated as follows:

$$DSC = \frac{2 \text{ x } TP}{2 \text{ x } TP+FP+FN} \qquad (5)$$

International Journal of Scientific & Engineering Research, Volume 11, Issue 5, May-2020
ISSN 2229-5518

Where by, **TP** represents the true positive rate, **FP** is the false positive rate, and **FN** is false negative rate.

4.3 Experimental Results

The proposed model in this study is compared with existing methods [23] under the same dataset, to validate the performance. Nested U-Net, Parallel Capsule Net, and the baseline methods that have been proven to show some improvement are used to compare with our proposed architecture. Table 1 gives the experimental results. Here, DSC is adopted as the criterion for comparison.

Table 1: Experimental results table

Model	*Dice (%)*
Nested U-Net	89
Parallel Capsule Net	90.2
Original 3D U-Net	88
Proposed Architecture (with guiding mechanism)	**90.5**

5 Conclusion

In this study, based on the results, it can be concluded that the original CNN architecture whereby the results were excellent. Still, on introducing guiding mechanism, the results became even much better. This trend suggests for data used to be more productive. Compared to traditional CNNs, the methods proposed were capable of equivalent or higher precision, thus substantially lowering the number of parameters to practice. Besides, speaking about the accomplishment of Deep Neural Network (DNN), One can't leave behind the achievements in the medical domain, and at times the performance is better than humans. At this age and time is without a doubt that Deep learning has brought significant impact to the lives of millions in different domains, medical and education are part of them. Innovative methods of integrating more than one method techniques were introduced, which has also proven to be successful. Techniques which involve using data-efficient models in exploring the potentials in solving complicated challenges with limited access of dataset is essential since it will bring about performance improvement in general.

Acknowledgment

The work is supported in part by Shanxi Province Natural Science Foundation Project (grant No.201801D121138) to author Li Fenglian. The author Fredy Malack Sabuni cannot express enough thanks to Prof. Li Fenglian for providing guidance and supervision for this research topic. Fredy Malack Sabuni and Ntikurako Guy Fernand Kazihise also sincerely acknowledge Taiyuan University of Technology for the opportunity presented to them. The authors also offer gratitude for reviewers who will take their time for this proposed research topic.

References

[1] Donnan G A , Fisher M , Macleod M R , et al. Stroke[J]. The Lancet, 2008, 371(9624):1612-1623.

[2] Baird AE, Benfield A, Schlaug G,等. Enlargement of human cerebral ischemic lesion volumes measured by diffusion-weighted magnetic resonance imaging.[J]. Annals of Neurology, 2010, 41(5):581-589.

[3] Hb V D W , Van G J . Clinical practice. Acute ischemic stroke[J]. New England Journal of Medicine, 2007, 357(6):572.

[4] Jacobs M A , Mitsias P , Soltanian-Zadeh H , et al. Multiparametric MRI Tissue Characterization in Clinical Stroke With Correlation to Clinical Outcome : Part 2[J]. Stroke, 2001, 32(4):950-957.

[5] Magnetic resonance imaging and computed tomography in emergency assessment of patients with suspected acute stroke: a prospective comparison," *The Lancet*, no. 369(9558):293–298, 2007.

[6] Maier O , Wilms M , Janina V D G , et al. Extra Tree forests for sub-acute ischemic stroke lesion segmentation in MR sequences[J]. Journal of Neuroscience Methods, 2015, 240:89-100.

[7] Maier O , Christoph Schröder, Forkert N D , et al. Classifiers for Ischemic Stroke Lesion Segmentation: A Comparison Study[J]. Plos One, 2015, 10.

[8] Moeskops P , Viergever M A , Adriënne M. Mendrik, et al. Automatic segmentation of MR brain images with a convolutional neural network[J]. IEEE Transactions on Medical Imaging, 2017, 35(5):1252-1261.

[9] Dolz J , Desrosiers C , Ayed I B . 3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study[J]. Neuroimage, 2017:S1053811917303324: 456-470.

[10] Ronneberger O , Fischer P , Brox T . U-Net: Convolutional Networks for Biomedical Image Segmentation[C]// International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer International Publishing, 2015.

[11] Kabir Y, Dojat M, Scherrer B, Forbes F, Garbay C. Multimodal MRI segmentation of ischemic stroke lesions. Conf Proc IEEE Eng Med Biol Soc.(2007) 2007:1595–8. doi: 10.1109/IEMBS.2007.4352610

[12] Maier, O., Wilms, M., von der Gablentz, J., Krämer, U., Handels, H., 2014. Ischemic stroke lesion segmentation in multi-spectral MR images with support vector machine classifiers. Medical Imaging 2014: Computer-Aided Diagnosis. International Society for Optics and Photonics, p. 903504.

[13] Maier, O., et al., 2017. ISLES 2015-A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. Medical Image Analysis 35, 250-269.

[14] S. Antol, A. Agrawal, J. Lu, M. Mitchell, D. Batra, C. L. Zitnick, and D. Parikh. Vqa: Visual question answering. *arXiv preprint arXiv:1505.00468*, 2015

[15] Y. Zhu, O. Groth, M. Bernstein, and L. Fei-Fei. Visual7w: Grounded question answering in images. In *CVPR*, 2016

[16] Y. Wang, Z. Deng, X. Hu, L. Zhu, X. Yang, X. Xu, P.-A. Heng, and D. Ni, "Deep attentional features for prostate segmentation in ultrasound," in *MICCAI*, 2018.

[17] D. Nie, Y. Gao, L. Wang, and D. Shen, "ASDNet: Attention based semi-supervised deep networks for medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 370–378.

[18] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert, "Attention gated networks: Learning to leverage salient regions in medical images," *Medical image analysis*, vol. 53, pp. 197– 207, 2019.

[19] ASDNet: D. Nie, Y. Gao, L. Wang, and D. Shen,"Attention based semi-supervised deep networks for medical image segmentation," *in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer,* p. 370–378. , 2018.

[20] Shyr A , Darrell T , Jordan M , et al. Supervised hierarchical Pitman-Yor process for natural scene segmentation[C]// Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011.

[21] D. Jin et al." Attention-based 3D Convolutional Network for Alzheimer's Disease Diagnosis and Biomarkers Exploration,," *IEEE*

International Journal of Scientific & Engineering Research, Volume 11, Issue 5, May-2020
ISSN 2229-5518

16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, pp. 1047-1051., Italy, 2019,.

[22] Oktay O , Schlemper J , Folgoc L L , et al. Attention U-Net: Learning Where to Look for the Pancreas[J]. 2018.

[23] MD Sharique, Bondi Uday Pundarikaksha , Pradeeba Sridar, Rama Krishnan R S, and Ramarathnam Krishnakumar et al. "Parallel Capsule Net for Ischemic Stroke Segmentation." June 6, 2019, Indian Institute of Technology Madras, Chennai Tamil Nadu 60003: https://doi.org/10.1101/66113

[24] Simard P , Steinkraus D , Platt J C . Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis[C]// International Conference on Document Analysis & Recognition. IEEE Computer Society, 2003.

[25] W. Y. P. B. Guan Q, "Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images," *a pilot study. J Cancer.,* no. doi:10.7150/jca.28769, 2019 Aug 27.

[26] Lecun Y , Bengio Y , Hinton G . Deep learning[J]. nature, 2015, 521(7553):436

[27] Chen L , Zhang H , Xiao J , et al. SCA-CNN: Spatial and Channel-Wise Attention in Convolutional Networks for Image Captioning[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 2017.

[28] Voets M , Mllersen K , Bongo L A . Replication study: Development and validation of deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs[J]. Plos One, 2018.