A NEW DEEP-NET ARCHITECTURE FOR ISCHEMIC STROKE LESION SEGMENTATION

Nesrine Jazzar[1] and Ali Douik[2]

[1]University of Sfax National Enginneering School of Sfax NoCCS Lab, Tunis, Tunisia
[2]National Enginneering School of Sousse, University of Sousse

ABSTRACT

Ischemic stroke, brain cells death due to a lack of oxygen, is a leading cause of long-term disability and death. Accurate diagnosis and timely intervention can effectively improve the blood supply of the ischemic stroke area and minimize brain damage. Recent studies have shown the potential to use magnetic resonance imaging (MRI) to provide contrast imaging to visualize and detect lesions. However, manual segmentation of the stroke lesion produced by MRI is a tedious and time-consuming task. Therefore, the automatic ischemic stroke lesion segmentation method may show excellent advantages. In this paper, we propose a novel deep learning method used to detect and localize brain ischemic stroke, a generalization encoder-decoder by modifying U-Net architecture.

We integrate multi-path architecture into both encoder and decoder blocks to captures different levels of the encoded state, which helps in more robust decision-making for stroke lesion segmentation. In bottleneck of the architecture, we applied dilated blocks to improve the underlying predictive capabilities. The proposed method has been tested on the publicly accessible web platform provided by the MICCAI Ischemic Stroke Lesion Segmentation (ISLES) challenge. The results demonstrate that the proposed method achieves a mean dice coefficient 0.91 of with the training and 0.84 with the testing data respectively.

KEYWORDS

Ischemic stroke segmentation, Convolutional neural network, U-Net, MRI, Dilated blocks.

1. INTRODUCTION

A cerebrovascular accident, i.e., a stroke is a failure of the blood flow that affects a large or small brain area. It occurs when a blood vessel is blocked or ruptured, and it causes the nerve cells to die, which are deprived of oxygen and essential nutrients for their appropriate function. The severity of the stroke depends on the location and the extent of the affected brain areas. According to the World Health Organization (WHO), an ischemic stroke is the leading national cause of acquired physical disability in adults and the second leading cause of death globally [1]. Early diagnosis and timely intervention are very critical for the recovery of stroke patients.

Magnetic resonance images (MRI) are the standard gold examination, they provide essential information for optimized treatment, eliminating a haemorrhagic accident, and detecting the lesion area from the first hour after the onset of clinical signs. MRI is much more accurate than a CT scan detecting multiple or small lesions or assessing the necrotic area's extent. These elements are essential for the patient's prognosis and the treatment decision.

David C. Wyld et al. (Eds): CMLA, CSEIT, NETCoM, NLPD, GRAPH-HOC, WiMoNE, CIoT, NCS - 2022
pp. 91-100, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121108

However, the automatic identification and segmentation of ischemic stroke lesions is not a trivial task because of the scarcity of datasets, image complexity, and the high variability of stroke's location contrast shape.

Most of the automatic segmentation methods use hand-designed features. Recently, there has been a growing interest in applying CNNs in image classification and segmentation. Recent studies show that it is more effective and suitable for complex neuroimaging tools such as neurological disorders and psychiatrists. Kaminatas et al. [2] proposed an approach for the segmentation of brain lesions using multimodal brain MRI based on 11-layer-deep, multi-scale 3D convolutional neural networks (CNNs) called Deep Medic. The proposed new training scheme is based on two main components: a 3D CNN, which produces exact flexible segmentation maps, and a fully connected 3D CRF conditional random field, which imposes regularization constraints on the CNN output and produces the segmentation labels. Chen et al. [3] proposed a framework with two CNN modules to segment stroke lesions using DWI in MRI. The first CNN was a combination of two DeconvNets (EDD Net), and the second one was a Multi-Scale Convolutional Label Assessment Network (MUSCLE Net) to focus on lesions detected at a small scale and aim to reduce false potentials detected by the EDD network. The dataset was constructed with clinically acquired DWI scans of 741 patients with acute stroke, exhibiting a high lesion detection rate and high accuracy. Liangliang Liu et al. [4] proposed a new Res-CNN automatic segmentation network that combines a similar U-shaped architecture with residual units. This network could alleviate the problem of the leakage gradient. The architecture of Res-CNN consists of 10 convolutional layers, 4 residual units, 4 concatenations layers, 4 deconvolution layers, and some batch normalization (BN) layers, and Leaky Rectified Linear Units (LReLU) [5]. The dataset was constructed with DWI scans and T2-to-DWI fusion (DWI-T2) as the multi-modality of input to improve lesion segmentation performance. Zhang et al. [6] proposed a fully convolutional and densely connected neural network (3D FC-DenseNet) to segment stroke lesions from DWI diffusion-weighted images. The network could use contextual information and learn end-to-end discriminating characteristics. The network is built based on the idea of densely connected convolutional networks, which allows each layer to take as input all of its previous feature maps, and two layers of a DenseNet network are directly connected. Liu et al. [7] proposed a Residual Structure Fully Convolutional Network (Res-FCN) to segment ischemic stroke lesions from multimodal MRI scans. In Res-FCN, the residual block can capture characteristics of large receptive fields for the network. Havaei et al. [8] proposed a CNN approach to segment subacute and ischemic stroke lesions from DWI, FLAIR, and T2 diffusion-weighted images. Lucas et al. [9] proposed a fully convolutional neural network (FCN) based on 2D-UNet networks with multiscale information propagation to segment acute stroke lesions. Some of the techniques give erroneous segmentation results when the lesions are small especially in the case of ischemic stroke segmentation.

The endeavor of this paper proposes a new deep learning architecture by modifying the U-Net [10] model to perform a fully automated stroke lesion segmentation task. We integrate multi-path architecture into both encoder and decoder blocks to outstanding feature representation ability and preserve low-level information. This multi-path architecture captures different levels of the encoded state, which helps in more robust decision-making [11] for stroke lesion segmentation. Furthermore, we applied dilated blocks in the bottleneck of the architecture to improve the underlying predictive capabilities [12]. We have conducted experiments on the publicly accessible web platform provided by the MICCAI Ischemic Stroke Lesion Segmentation (ISLES) challenge [13] with different images modalities.

This paper is organized as follows. Section II provides an overview of the dataset utilized for stroke segmentation. Section III illustrates in detail the proposed deep method. Then, Section IV includes the evaluation metric used to evaluate the segmentation results, followed by the findings of the experimental results in section V. Finally, the efficiency of the proposed architecture is summarized in section VI.

2. DATASET DESCRIPTION

2.1. Data Acquisition

The proposed model was assessed on a multi-modal MRI sub-task, the sub-acute ischemic stroke segmentation (SISS) of the MICCAI ISLES 2015 challenge [13] dataset. The SISS dataset contains 64 cases (28 training and 36 test) of patients with sub-acute ischemic stroke. Each patient has four co-registered MRI modalities, namely Fluid-Attenuated Inversion Recovery (Flair), Diffusion-Weighted Imaging (DWI), T1, and T2. The images acquired had a 3D voxel size of 230 ×230 ×153 with an image resolution of 1mm × 1mm × 1mm spacing. The dataset is provided with pixel-accurate ground truth labels. The lesions in this dataset were small and diffuse.

2.2. Data Pre-processing

The training data is one of the essential keys to obtaining a high prediction model's accuracy. That is why pre-processing is mandatory. Hence, we have prepared data for use in the following steps: First, we eliminated any black slices with no data for every patient. Second, all the images of 230×230×3 shape are timed to a new shape of 129×129×3, which is more convenient for the following work steps.

The total number of images in our dataset was 3900 images. Therefore, the data was further augmented using randomly a left-right flip of the images, horizontal flipping, shearing, rotation, and zooming to produce six times as much data, giving 19000 total images. Later, the intensity values of these images are normalized in the range of [0,1] based on the minimal value. The main reason for pre-processing is to increase the robustness and validation accuracy of the network and tackle data insufficiency issues in medical imaging. The final step is data partitioning: 80% for the training and the rest 20% is for the validation.

3. METHODOLOGY

3.1. Network Description

This work proposes a deep learning architecture using a residual CNN inspired by the U-Net [10] architecture multipath network and dilated convolution, illustrated in Figure 1. Our architecture is an encoder-decoder network that takes advantage of a multipath network by integrating M-blocks in a single path as the elementary module [11]. Using a multipath procedure will enhance the possibility to constructs a more extended feature than a single path. Both the encoder and decoder path comprise several M-Blocks, as shown in Figure 2. The M-Block consists of 4 different paths Pi where i in 0 to 3. In each path, we use a different number of convolutional layers. For example, P1, P2, and P3 have one, two, and three convolutional layers, respectively. This technique will make the learning more comprehensive features easier and more precise than a single path by allowing flexibility to the amount of encoding/decoding required for precise segmentation. On the other side, to get advantages of residual connection and minimize information loss along with the depth of the network, P0 does not include any convolutional layer. The convolutional layers consist of Kernel size 3 × 3 number of filter zero-padding followed by batch-normalization and ReLU activation [14]. The outputs of the four paths are concatenated and passed through the activation function. Each encoder block output is processed in two different ways. The output is max-pooled and sent to the next encoder block as its input, or the same output is concatenated with the transposed output of the lower encoder block and sent to the corresponding decoder block as feedback which enhances the feature. In the encoder block, the max-pooling is done 5 times on the image data, resulting in 6 times smaller than the initial state. The decoder is the same as the encoder, except the max-pooling operation is replaced by a convolutional transpose operation. The output of the last decoder block is passed through a convolutional block with a single filter of kernel size 1 × 1 with sigmoid ($\frac{1}{1+e^{-x}}$) activation function.

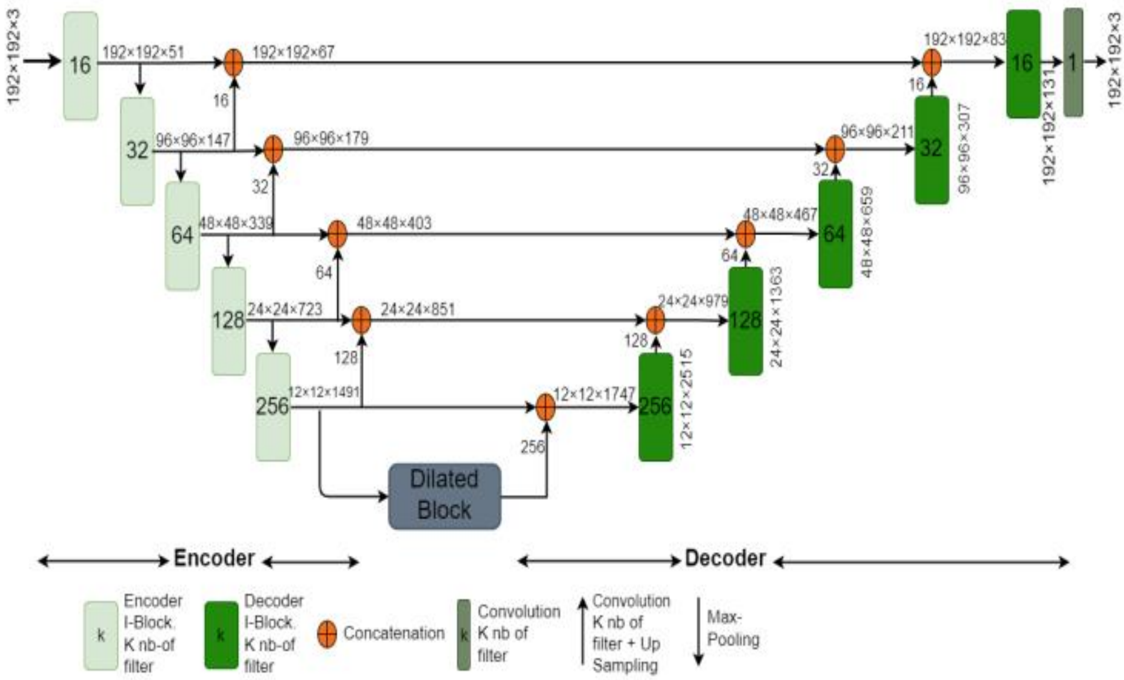

Figure 1. The architecture of the proposed model for ischemic stroke segmentation

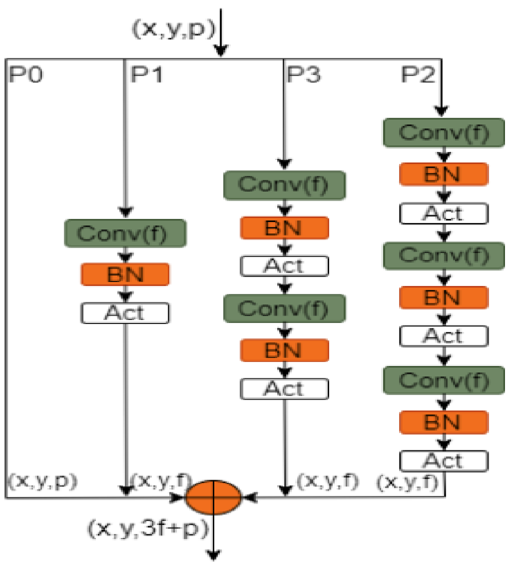

Figure 2. Schematic diagram of the M-Block

In our network, we also integrate the dilated block at the lowest layer of the network to summarize the global information and generate the output of the encoder. Dilated convolution also named Atrous convolution, was originally developed for Wavelet decomposition. The goal of using of a dilation convolution is to insert a hole between the pixels in the convolutional kernel to capture the texture information with different receptive fields. The receptive field is how large the pixels in the high-level feature map are affected by the original image of each layer of the convolutional neural network. We use low dilation rate convolutional to capture the texture information on a small scale and use a high dilation rate convolutional to capture the texture information on a large scale. The dilated block illustrated in Figure 3, consisting of four dilated convolution layers employed in the bottleneck of the network, is configured such that the first layer uses a dilation rate of one. Furthermore, each subsequent layer increases the dilation rate by a multiple of two as proposed in [13]. The dilation rate is 1, 2, 4, and 8. Each convolutional kernel's feature scale is $(2k+1)^2$, where k is the kernel's dilation rate. However, the features extracted from the dilated convolution result produce a different scale of 3 × 3, 5 × 5, 9 × 9, and 17 × 17 as shown in Figure 4. We applied Batch normalization for the output of each dilated convolution layer to enhance the network's stability, followed by ReLU is used as an activation function. Then, a concatenation of the four ReLU outputs, followed by 1 × 1 convolution, to reduce the dimension, Batch normalization, and finally the ReLU activation function. We chose ADAM optimizer with an initial learning rate of Lri = 10^{-4}, decay factor =0.2, step =2.

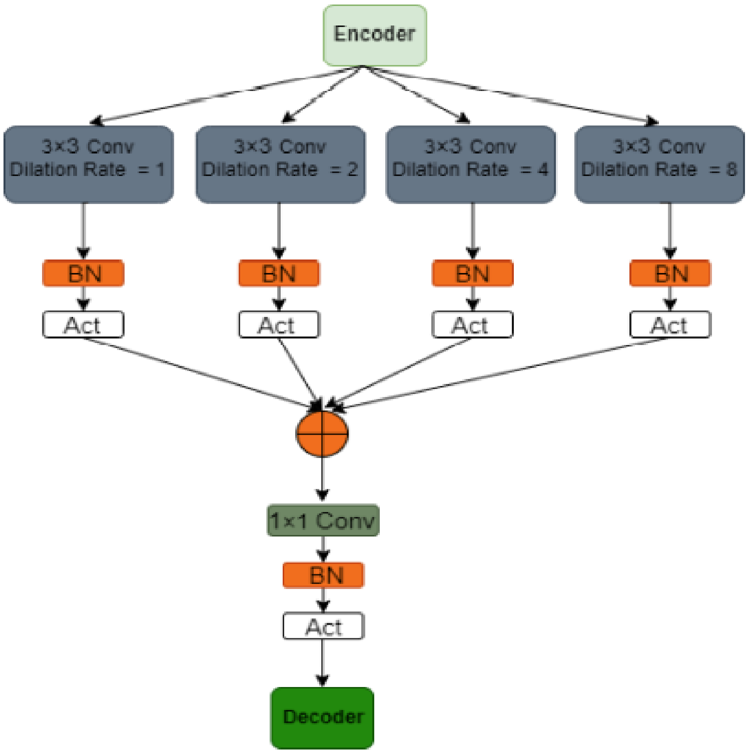

Figure 3. Schematic diagram of the dilated Block

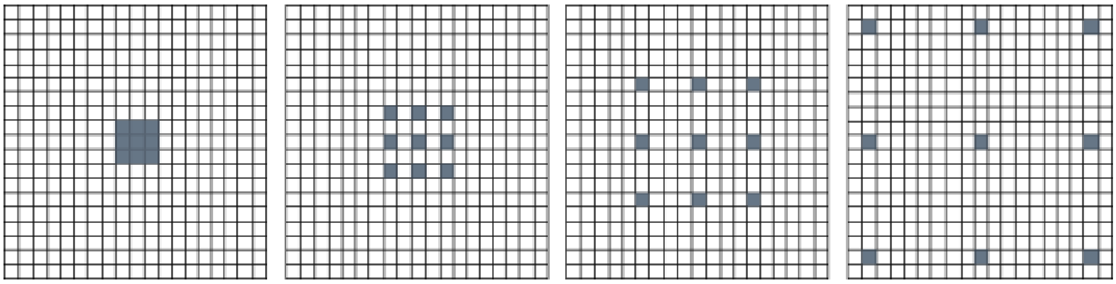

Figure 4. Dilated convolution with 3×3 Kernel with a dilation rate of 1, 2, 4 and 8

3.2. Training Techniques

The input data dimension is 192×192×3, and the segmented result obtained from the network is 192 × 192 × 1. In the training process, the weight initialization plays a vital role in converging the model. It is initialized using he-normal, which draws the ample from a truncated normal distribution centred on 0 with $\sigma = \sqrt{\frac{2}{n}}$, where n is the number of input units in the weight is updated using the Adam optimizer a batch size of 5. The loss function (Binary Cross Entropy BCE, Dice Loss, BCE Dice Loss, Sigmoid BCE) is used to measure the error, as defined in Table 1. It is also the function to be minimized during training, and we use early stopping by monitoring validation loss function with a patient of 10 to bypass the severe class imbalance problem. Our proposed model has trained these four-loss functions separately, as mentioned in Table 2.

Table 1. Loss Function.

Binary Cross Entropy(BCE)	-y $\log(\hat{y}) + (1-y)(\log(1-\hat{y}))$
Dice Loss	1 - 2 $\left\lvert\frac{y\cap\hat{y}}{y\cap\hat{y}}\right\rvert$
BCE Dice Loss	BCE-2 $\left\lvert\frac{y\cap\hat{y}}{y\cap\hat{y}}\right\rvert$
Sigmoid BCE	-y $\log(f(\hat{y})) + (1-y)(\log(1-f(\hat{y})))$

Table 2. Performance of proposed model in term of Dice coefficient on each modality on SISS Dataset for various loss functions.

Modality	BCE		Dice Loss		BCE Dice Loss		SCE	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
DWI	0.871	0.814	0.892	0.809	0.855	0.714	0.685	0.581
Flair	0.895	0.827	0.916	0.844	0.881	0.759	0.809	0.772
T1	0.784	0.671	0.858	0.699	0.877	0.738	0.815	0.793
T2	0.755	0.674	0.863	0.695	0.869	0.740	0.694	0.618

4. Evaluation Metrics

Evaluation metrics are essential tools to analyse the performance of segmentation. The value of each criterion increases with the quality of the segmentation result. These values have been standardized to facilitate their comparisons. A criterion value close to 1 reflects an excellent segmentation result. Our model has been evaluated on four widely used quality metrics, which are defined as follows.

4.1. Dice Similarity Coefficient

The Dice score (DSC) [16] measures the similarity; overlap between the manually segmented ground truth and our segmentation results, which will be calculated as follows:

$$DSC = \frac{2TP}{2TP + FP + FN}$$

Where, True Positive (TP) indicates that the method correctly segmented pixels. False Positive (FP) indicates the pixel that the method classifies negative as positive. False Negative (FN) denotes that the positive pixel is incorrectly classified as negative by the segmentation method.

4.2. Accuracy

The accuracy related to the percentage of pixels that were correctly predicted over the total number of pixels segmented in an image [19]. It is defined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

4.3. Specificity

The specificity measures the percentage of pixels correctly predicted as belonging to the background region among all the pixels belonging to the background. It is defined as:

$$Specificity = \frac{TP}{TP + FP}$$

4.4. Sensitivity

The sensitivity measures the percentage of pixels correctly segmented that are correctly identified. It is defined as:

$$Sensitivity = \frac{TP}{TP + FN}$$

5. Results and Discussion

The proposed architecture is trained in ubuntu environment, Dell Predator, inter-core i5, 24 GB RAM installed with NVIDIA GeForce GTX 1050 Ti. Keras and TensorFlow are used as frameworks to implement the model. The detailed statistics of the outcome are shown in Table 2, 3. We have analysed that the Dice Loss function on the SISS dataset gives the optimum results for all modalities, which is better than other functions on both training and testing datasets. However, the Dice loss function is the most suitable loss function for all modalities. We also noticed that the FLAIR and DWI are the most suitable image modalities while producing consistent outcomes using the Dice loss function. Finally, the average has been calculated to compare the performance of our architecture with some of the well-known methods which are shown in Table 3. It should be noted that our model surpasses the state-of-the-art results marginally. The visual performance of our model on the SISS dataset is shown in Figure 5, which is a curve for DSC and accuracy for training and validation set, and Figure 6.

Figure 5. From left to right: Plot for DSC and accuracy for training and validation set

Table 3. Comparison of Dice for various methods on the SISS training and testing datasets. The average is calculated over the total number of patients. The showcased data has been obtained from the ISLES 2015 Challenge.

Method	DSC	
	Training	Testing
Robben et al. [18]	057±0.28	0.43 ±0.30
Maier et al. [12]	0.58 ±0.29	0.42 ±0.33
Feng et al. [19]	0.63 ±0.28	0.55 ±0.30
Chen et al. [3]	0.55 ±0.29	0.44 ±0.30
proposed	0.91 ±0.11	0.84 ±0.24

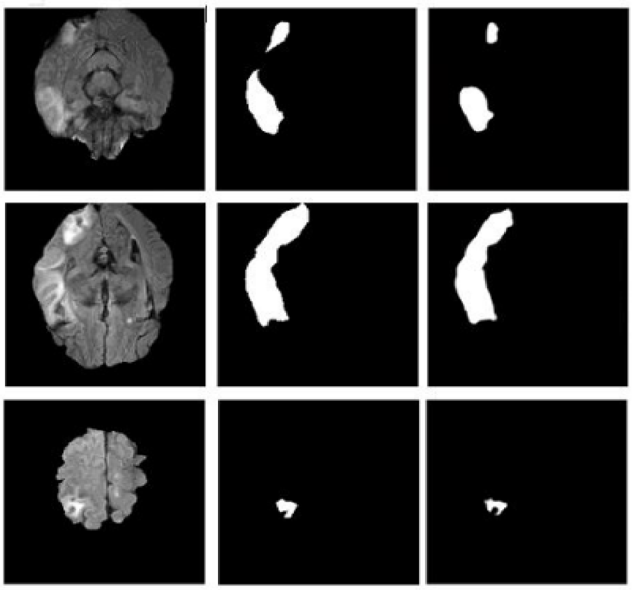

Figure 6. Visual result of our model from the axial view for three different patients on SISS dataset: Original image, ground truth and predicted image

6. CONCLUSIONS

In this paper, we propose a fully automated ischemic stroke segmentation from various modalities inspired by U-Net architecture fused with a multipath network and we integrate a dilated block in the bottleneck. We optimized our model with various loss functions to tackle the severe class imbalance problem. However, we conclude that the Dice loss function gives the optimum results for all modalities. We also evaluate our architecture on a public challenge dataset SISS 2015, where its effectiveness and generalization capability are further demonstrated. However, the proposed architecture presents high segmentation accuracy with different modalities MRI images with a Dice coefficient for subsequent work, we aim to do a fusion of two different modalities such as FLAIR and DWI images such input for our model.

REFERENCES

[1] Global Diffusion of eHealth, (2017), Making Universal Health Coverage Achievable, Report of the Third Global Survey on eHealth, World Health Org., Geneva, Switzerland.

[2] Kamnitsas K, Ledig C, Newcombe VFH, Simpson JP, Kane AD, Menon DK, et al. (2017), Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation, Med Image Anal.

[3] Chen L, Bentley P, Rueckert D, (2017), Fully automatic acute ischemic lesion segmentation in DWI using convolutional neural networks, NeuroImage.

[4] Liangliang Liu, Shaowu Chen, Fuhao Zhang, Fang-Xiang Wu, Yi Pan, Jianxin Wang, (2019), Deep convolutional neural network for automatically segmenting acute ischemic stroke lesion in multi-modality MRI, part of Springer Nature.

[5] Nair V, Hinton GE, (2010), Rectified linear units improve restricted boltzmann machines, In: International conference on international conference on machine learning, pp 807–814,

[6] Zhang R, Zhao L, Lou W, Abrigo JM, Mok VC, Chu WC, Wang D, Shi L, (2018), Automatic segmentation of acute ischemic stroke from dwi using 3d fully convolutional densenets, IEEE Trans Med Imaging 37(9):2149–2160.

[7] Liu Z, Cao C, Ding S, Han T, Wu H, Liu S, (2018), Towards clinical diagnosis, automated stroke lesion segmentation on multimodal mr image using convolutional neural network.

[8] Havaei M, Dutil F, Pal C, Larochelle H, Jodoin PM, (2015), A convolutional neural network approach to brain tumor segmentation, In: International Workshop on Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries, pp 195–208.

[9] Christian Lucas, André Kemmling, Amir Madany Mamlouk, Mattias P. Heinrich, (2018), Multi-Scale neural network for automatic segmentation of ischemic strokes on acute perfusion images, IEEE 15th International Symposium on Biomedical Imaging (ISBI), Washington, D.C., USA.

[10] O. Ronneberger, P. Fischer and T. Brox, (2015), U-net: Convolutional networks for biomedical image segmentation, International Conference on Medical image computing and computer assisted intervention, pp. 234-241.

[11] S. Zagoruyko, A. Lerer, T.-Y. Lin, P. O. Pinheiro, S. Gross, S. Chintala, and P. Dollar, (2016), A multipath network for object detection.

[12] O. Maier, B. H. Menze, J. von der Gablentz, L. Häni, M. P. Heinrich, M. Liebrand, S. Winzeck, A. Basit, P. Bentley, L. Chen et al, (2017), Isles 2015-a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral mri, Medical image analysis, vol. 35, pp. 250-269.

[13] Vesal, S., Ravikumar, N., Maier, (2018), A.: Dilated convolutions in neural networks for left atrial segmentation in 3d gadolinium enhanced-mri.

[14] V. Nair and G. E. Hinton, (2010), Rectified linear units improve restricted boltzmann machines, in Proceedings of the 27th international conference on machine learning (ICML-10).

[15] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, (2017), Rethinking atrous convolution for semantic image segmentation.

[16] L. R. Dice, (1945), Measures of the amount of ecologic association between species, Ecology, vol. 26, no. 3, pp. 297–302.

[17] Chang, P.D, (2016), Fully convolutional neural networks with hyperlocal features for brain tumor segmentation, MICCAI- BRATS 2016 Multimodal Brain Tumor Image Segmentation Benchmark.

[18] D. Robben, D. Christiaens, J. R. Rangarajan, J. Gelderblom, P. Joris, F. Maes, and P. Suetens, (2015), A voxel-wise, cascaded classification approach to ischemic stroke lesion segmentation, in BrainLes 2015. Springer, pp. 254–265.

[19] C. Feng, D. Zhao, and M. Huang, (2015), Segmentation of Stroke Lesions in Multi-spectral MR Images Using Bias Correction Embedded FCM and Three Phase Level Set.

AUTHORS

Nesrine Jazzar received the B.A. degree in mathematical from Ali Bourguiba College, Monastir, Tunisia, in 2012, and obtained licence degree in computer science in 2015 and in 2017 she had the Master degree Industrial computer science. Since 2019, She is a Ph.D. student at the National Engineering School of Sfax, Tunisia. She is a member of the Networked Objects, Control, and Communication Systems (NOCCS) laboratory. Her research focuses on computer sciences, in particular, medical image segmentation.

ALi Douik Director at National Engineering School of Sousse. Having a doctorate in Electrical Engineering and a qualification to direct research. He was able to provide lessons in Automatics, system control, Image processing. He supervised several doctoral theses. He published numerous scientific articles in peer-reviewed international conferences and journals.