1176: ARTIFICIAL INTELLIGENCE AND DEEP LEARNING FOR BIO-MEDICAL APPLICATIONS



# Deep learning based neural network application for automatic ultrasonic computed tomographic bone image segmentation

Fradi Marwa<sup>1,2</sup> • El-hadi Zahzah<sup>2</sup> • Kais Bouallegue<sup>3</sup> • Mohsen Machhout<sup>1</sup>

Received: 21 July 2020 / Revised: 30 September 2021 / Accepted: 17 January 2022 / Published online: 16 February 2022 © The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

Deep-learning techniques have led to technological progress in the area of medical imaging segmentation especially in the ultrasound domain. In this paper, the main goal of this study is to optimize a deep-learning-based neural network architecture for automatic segmentation in Ultrasonic Computed Tomography (USCT) bone images in a short time process. The proposed method is based on an end to end neural network architecture. First, the novelty is shown by the improvement of Variable Structure Model of Neuron (VSMN), which is trained for both USCT noise removal and dataset augmentation. Second, a VGG-SegNet neural network architecture is trained and tested on new USCT images not seen before for automatic bone segmentation. Therefore, we offer a free USCT dataset. In addition, the proposed model is implemented on both the CPU and the GPU, hence overcoming previous works by a value of 97.38% and 96% for training and validation and achieving high segmentation accuracy for testing with a small error of 0.006, in a short time process. The suggested method demonstrates its ability to augment USCT data and then to automatically segment USCT bone structures achieving excellent accuracy outperforming the state of the art.

Keywords VSMN-VGG-SegNet · USCT · Segmentation · Time process · GPU

Fradi Marwa marwafradi32@gmail.com

<sup>&</sup>lt;sup>1</sup> Physic Department of Faculty of Sciences of Monastir, Monastir University, Monastir, Tunisia

<sup>&</sup>lt;sup>2</sup> Laboratory of Informatics, Image and Interaction (L3i, France), La Rochelle University, La Rochelle, France

<sup>&</sup>lt;sup>3</sup> ISSAT Sousse, Sousse University, Sousse, Tunisia

### 1 Introduction

For a long time, deep-learning-based neural network systems have been inspired by biological observations [3]. These systems have been developed for the resolution of control and recognition problems of certain characteristics in an image. Recently, deep learning has revolutionized the biomedical field. Accordingly, it has managed to become very popular in the field of image processing, more specifically in the field of medical imaging [8], involving MRI, X-Rays, CT and ultrasonic images [7, 23]. Accordingly, it can extract extraction of different anatomical structures, to ensure the automatic segmentation of the regions of interest [24]. Nowadays, the Ultrasonic Computed Tomography (USCT) device, an existing new technique, has revolutionized X-rays and ultrasonic imaging [28]. It is a non-invasive and non-ionizing technique. However, USCT images are noisy and difficult to analyze, given the inhomogeneity of pixels and the high frequency of transmitted ultrasound waves [9, 24]. Indeed, the idea of USCT medical image analysis using deep-learning-based neural network techniques has remained a hot topic of interest in the field of USCT medical imaging [30]. In this context, we put forward a deep learning model to ensure the automatic segmentation of bone USCT images. To overcome this issue, processing analysis based on deep learning techniques for USCT images is proposed. Such processing comprises segmentation to detect the bone boundaries and extract the characteristics of each bone region. Each ultrasonic tomographic image has three layers of bone structures to automatically segment USCT bone images, such as the cortical bone, the cancellous bone and the medullary cavity. The detection of these three structures from a noisy USCT image is very difficult and remains a problem to overcome. Above all, it is necessary to eliminate the background which presents a big noise, to assist clinicians in determining the diagnosis in USCT images of bones such as fractures, osteoporosis and tumors.

Our work aims to carry out Convolutional Neural Network (CNN) learning with VGG-SegNet and VGG-Unet models applied on a USCT image dataset, in order to achieve an automatic segmentation of the region. Thus, we improve a Variable Structure Model of neurons (VSMN) [3] and apply it on medical images to get a significant increase in data, given the problem of unavailability of USCT images [28].

The rest of the paper is partitioned as follows. Section 2 introduces the state of the art. Section 3 presents the experiments and the methodology. Section 4 provides the achieved results. The discussions and the conclusion are respectively given in section 5 and section 6.

The contribution of this paper is presented as follows: We have provided original USCT data set available for free on (https://www.kaggle.com/fradimarwa/usct-dataset-of-bones), for USCT researchers. Then we have made the design of a new neural network system for USCT images segmentation. First, we design a new variable structure of neural network called VSMN for USCT image processing. Then, we optimize the VGG-SegNet network, to automatically segment USCT images. Thus, our work presents the first study, using an end to end (VSMN-VGG-SegNet) neural network application based on deep learning to automatically segment USCT images of bones. Indeed, the segmentation of USCT images of bones has not been explored in the literature using deep learning, given the difficulty of obtaining a large amount of USCT data [28]. Finally, our proposed system can be applied to any database, such as a real scene image database and implemented on GPU. Finally, we have achieved promising accuracies and a short time process compared to the state of the art.

## 2 State of the art

Classical approaches for ultarasonic medical image segmentation have employed machine learning techniques [26, 32]. These techniques include the Atlas model and the dictionary learning. The Atlas model has been developed for the segmentation of medical images, but the limits have remained noisy. It was applied on MRI tomographic images to detect lung tumors in [15] and simultaneously improve the quality of MRI images. In [9], the wavelet transforms yielded excellent results in USCT image analysis. Furthermore, a propounded method using the K-means, and the Ostu method yielded the best performance in USCT image segmentation and led to automatic diagnosis detection in [10]. In [12, 31], machine learning for ultrasound image segmentation proved its excellence with promising accuracy results. However, the machine learning technique applied to USCT breast images demonstrated its ability to achieve excellent segmentation results, as presented by the authors in [12]. This method was based on semi-automated 3D segmentation through the detection of the breast boundary in coronal slice images. In [6], the active contour method was massively used in the segmentation of ultrasonic image processing. It was used to avoid the noise in USCT image. This method was applied by Lasaygues on a tomographic image made with the USCT of a paired bone, but the results were not satisfactory and the detection of the distances between the two bone forms (tibia and fibula) were not possible considering the noise present in the image [14].

These machine learning segmentation techniques commonly used in the past have been less effective than deep learning counter parts because they have used rigid algorithms and required human intervention and expertise.

However, modern ultrasound image analysis techniques rely on deep-learning technologies [34] where the segmentation of ultrasound medical images is a topic of interest in the field of medical imaging. Indeed, deep learning is known as a process that allows computational models composed of multiple processing layers to learn representations of data with multiple levels of abstraction [21], for the automatic segmentation of different anatomical structures. It involves automatic segmentation methods previously classified as supervised or unsupervised [35]. For the supervised methods, segmentation requires the operator interaction throughout the segmentation process, while the unsupervised methods generally require the operator intervention only after the end of the segmentation process. The unsupervised methods are preferable to ensure a reproducible result [35]. However, the operator interaction is still necessary for error correction in the event of failure at the result level. The application of deep-learning-based neural networks, such as the Convolutional Neural Network (CNN), SegNet, Unet and X-Net, has improved the USCT image segmentation. Indeed, the segmentation of medical images based on the CNN, known as multilayer neural networks specializing in shape recognition tasks [18], relies on several deep neuron networks alternating between the application layers of convolutions and max pooling. It has been adapted to the hand and brain segmentation [35]. In [30], authors implemented the CNN and the Convolutional Long Short-Term Memory (ConvLSTM)-based deep learning models for Covid-19 class detection, achieved results demonstrated excellent accuracies. In [4], XNet was proposed for X-ray image segmentation while having producing accuracy of 92% and AUC of 98%. These results surpassed the conventional treatment of medical images. However, SegNet was used for image labelling. It only depended on the fully learned function to get the label prediction. Furthermore, Unet achieved 93% accuracy by detecting different human bones and skeletons [18]. In addition, deep learning was applied to tomographic MRI images for the detection of lung tumors in [15] while improving the MRI image quality. In [19], deep learning was used to combine a neural GNN network and Unet to perform the automatic segmentation of the airways in the rib cage. Deep-learning-based Unet for bone structure segmentation in CT X-ray tomographic images presented very promising results. It showed its efficiency in automatically segmenting the bone structures of the femur in MRI images [2]. It also helped clinicians to determine the diagnosis [11] by ensuring the automatic segmentation of the intervertebral disc, while achieving segmentation precision with a value of 83%.

# 3 Methods

# 3.1 Experimental method

Our experiments are done using a new prototype, called USCT, providing a new technique for bone imaging, which has revolutionized X-rays, MRI and ultrasound techniques [28]. The used device is an ultrasonic scanner, consisting of a 2D-circular antenna with 8 transducters distributed over 360° every 45°. Accordingly, the eight transducters are piezo-composite elements whose frequencies are 1–3 MHz as depicted in Fig. 1 and detailed in [9]. In addition, the imaging process gives us 50 USCT bone images which will be increased in the following section due to our proposed method.

# 3.2 Synoptic flow of proposed method

The suggested structure remains a hybrid model involving an optimized VSMN [3] and a neural VGG-SegNet network. Our proposed neural network architecture is presented, as depicted in Fig. 2. Our approach aims to optimize the VSMN, by modifying the activation function and then making it accessible for a medical image processing application, performing an optimal number of filtered USCT images. These images obtained by the VSMN, serve as an



Fig. 1 Ultrasonic Computed Tomography device



Fig. 2 Synoptic flow of proposed method

input to a second neuron network, called the neural VGG-SegNet network, which ensures the automatic segmentation with background removal.

#### 3.2.1 VSMN model

**Mathematical theorems** A neural network model was developed by [3], called the VSMN. This model is introduced by the following equations. The VSMN structure needs four variables (n, p, q, k) where n and q are related to the model behavior, p is related to the threshold position of the model, k represents the neuron's polarity,  $\tau$  represents the constant of time, p and q are real numbers and  $\alpha$ ,  $\beta$  and  $\lambda$  are positive real numbers.

$$\dot{\mathbf{u}} = \frac{-(\mathbf{u} + \mathbf{p})}{\tau} + (\mathbf{u} + \mathbf{p})\mathbf{v} \ \mathbf{f}(\mathcal{B}\mathbf{v})\mathbf{f} \ (\lambda \ (\mathbf{u} + \mathbf{p}))$$
(1)

$$\dot{\upsilon} = -\alpha v + k(u+q)^2 + \alpha f^2(\lambda(u+p))$$
<sup>(2)</sup>

$$g(u) = v + \alpha v = k(u+q)^n + \alpha f^2(\lambda(u+p))$$
(3)

Compared to the model studied by [3], a modification in the activation function is made in our VSMN neural network, as shown in Eq. (4). It is satisfactory to focus on function g(x) described by Eq. (5).

$$f(t) = \exp^{t} \text{ where } t = -x + p \tag{4}$$

 $g(x)=k[(\neg x+q)$ 

Ola approach is to optimize the VSMN, by modifying the activation function and then making it accessible for a medical image processing application performing optimal filtering of USCT images, hence the automatic increase in images. From Eq. (5), we get the following equations:

$$Z = (-x + q)^{n}, \quad Y = \exp(-x + p)^{2}, g'(x) = (Z^{*}Y)'$$
(6)

$$\mathbf{h}(\mathbf{x}) = \mathbf{g}'(\mathbf{x}) \tag{7}$$

$$\mathbf{h}(\mathbf{x}) = \left[-n(-x+q)^{n-1}\right)\left]e^{(-x+p)^2}\right] + (-x+q)^n \left[\left[-2(-x+p)\right]e^{(-x+p)^2}\right]$$
(8)

$$= \left[-e^{(-x+p)^{2}}(-x+q)^{n-1}\right]\left[n+2(-x+p)(-x+q)\right]$$

Function h (x) has three solutions:

$$\left[e^{(-x+p)^{2}}(-x+q)^{n-1}\right] = 0; x1 = q$$
(9)

$$[n+2(x-p)(-x+q)] = 0; x^2 = \frac{-b+\sqrt{b^2-4ac}}{2a} = \frac{(p+q)+\sqrt{(-p-q)^2+2n}}{2} \quad (10)$$

$$[n+2(x-p)(-x+q)] = 0; x3 = \frac{-b-\sqrt{b^2-4ac}}{2a} = \frac{(p+q)-\sqrt{(-p-q)^2+2n}}{2}$$
(11)

**VSMN architecture** The VSMN model is produced as a cascade architecture. The output of the first neuron is considered the input of the second neuron for each layer. Indeed, k represents the polarity of neurons. It can be with positive or negative polarity. Actually, n represents the number of layers and p and q the parameters for each neuron. The model of the neuron architecture is shown in Fig. 3 with positive polarization. Indeed, the use of a negative polarity k = -1, gives USCT images with poor quality. For this reason, the choice of positive polarity is done. The internal architecture of our VSMN model is described by Fig. 4 and its mathematical analysis is more detailed in Tables 1 and 2.

#### Fig. 3 Neuron Model



### 3.3 VSMN implementation on USCT images

#### 1rstCase:

Starting by the first layer, for n = 0 and from Eq. (5), we get g(x) as described by Eq. (12).

For 
$$n = 0, p = q = 1, k = 1$$
  
 $g(x) = \exp(-x + p)^2$ 
(12)

The g(x) curve as depicted in Fig. 5 describes the VSMN behavior in the first layer where n = 0.



Fig. 4 VSMN architecture (seven Layers L)

Layers	Parameters: n, p, q, k	$g(x) = k[(-x+q)^n]e^{(-x+p)^2}$ (5)
L1	n=0, p=q, k=1	$g(x) = exp(-x+p)^2$
L2	n=1, p=q, k=1	$g(x) = (-x+q)^1 exp(-x+p)^2$
L3	n=2, p=q, k=1	$g(x) = (-x+q)^2 \exp(-x+p)^2$
L4	n=3, p=q, k=1	$g(x) = (-x+q)^3 exp(-x+p)^2$
L5	n=4, p=q, k=1	$g(x) = (-x+q)^4 exp(-x+p)^2$
L6	n=5, p=q, k=1	$g(x) = (-x+q)^5 \exp(-x+p)^2$
L7	n=6, p=q, k=1	$g(x) = (-x+q)^6 exp(-x+p)^2$

Table 1 Mathmetical analysis of SVMN architecture in Fig. 4

The g(x) curve as depicted in Fig. 5 provide a deceasing behavior and then an increasing behavior. The SVMN behavior is explained by the following mathematical analysis equations:

For  $g(x) = \exp((-x + p)^2)$ , g(x) is a symetric function and x = 1 represent the axe of symetry.

$$\lim_{x \to -\infty} g(x) = \lim_{x \to +\infty} g(x) = +\infty$$
(13)

$$\lim_{x \to -2} g(x) = (\exp)9 \tag{14}$$

$$\lim_{x \to 1} g(x) = (\exp)0 = 1 \tag{15}$$

Equations (13), (14) and (15) shows that the VSMN curve has a decreasing behavior then an increasing behavior, where x = 1 represent the axe of symetry. The VSMN behavior has a great impact on the image quality as depicted in Fig. 6 and more explained by the following Eq. (16), (17),(18) and (19).

The implementation of our optimized VSMN proves its ability to remove noise from USCT images in a first step, and then to augment the number of USCT images, which is a hard task to achieve. Figure 6 shows USCT images in the first layer when n = 0. To conclude, the deeper we go, the better the quality of USCT images.

$$(a1): Y0 = \exp(-x + p)^{2}; n = 0, q = 1; p = 1$$
(16)

Layers	Parameters: n, p, q, k	$g(x)=k[\left(-x+q\right)^n]e^{\left(-x+p\right)^2}$	Y=Output
L1	n=0, p=q, k=1	$g(x) = exp(-x+p)^2$	$Y0 = exp(-x+p)^{2}$ Y1 = exp(-Y0+1)^{2} Y2 = exp(-Y1+1)^{2} Y3 = exp(-Y2+1)^{2}

Table 2 Mathematical analysis via internal architecture of layer



Fig. 5 VSMN with stable behavior for n = 0 and k = 1

$$(b1): Y1 = exp(-Y0 + 1)^2; n = 0, q = 1; p = 1$$
 (17)

$$(c1): Y2 = exp(-Y1 + 1)^2; n = 0, q = 1; p = 1$$
 (18)

$$(d1): Y3 = \exp(-Y2 + 1)^2; n = 0, q = 1; p = 1$$
(19)

#### 2nd case:

For the second layer, n = 1, g(x) is illustrated by the Eq. (20) and the VSMN behavior is depicted in Fig. 7(a). Moreover its robustness is shown to be implemented in USCT images as provided in Fig. 7(b).

The g(x) curve provided in Fig. 7 shows a deceasing behavior which is explained by the following mathematical analysis equations:

$$g(x) = k(-x+q)^{n} exp(-x+p)^{2}$$
  
For n = 1, p = q = 1, k = 1, g(x) = (-x+1)^{1} exp(-x+1)^{2} (20)



**Fig. 6** USCT output through layer 0 VSMN with stable behavior for n = 0 and k = 1, (a): Patella adult bone used to be imaged by USCT (b): USCT Results via layer 0



Fig. 7 VSMN behavior, n = 1, (a) VSMN curve, (b) output USCT image

For 
$$g(x) = (-x+1)^1 \exp(-x+1)^2$$
  

$$\lim_{x \to -\infty} \exp(-x+1)^2 = +\infty$$
(21)

$$\lim_{x \to -\infty} (-x+1)^1 = +\infty \tag{22}$$

From Eqs. (21) and (22) we get the Eq. (23) as follows:

$$\lim_{x \to \infty} g(x) = +\infty \tag{23}$$

The g(x) function is illustrated by Eq. (24) when  $x \rightarrow 1$ 

$$\lim_{x \to 1} g(x) = 0 \tag{24}$$

$$\lim_{x \to +\infty} g(x) = -\infty \tag{25}$$

$$\lim_{x \to +\infty} \left( \frac{g(x)}{x} \right) = -\infty$$
 (26)

Equations (23), (24), (25) and (26) have proved the g(x) behavior as depicted in Fig. 7, where the g(x) function converge to 0 through the second layer and the amplitude of the signal increases compared to which is in the first layer. This SVMN behavior has a great impact on the USCT image quality, as shown in Fig. 7(b) more parameter n increases, more the quality of image is going better. Thus, deeper we go in the SVMN, more higher the quality of images.

### 3<sup>rd</sup>Case:

For the third layer, for n = 2 and from Eq. (5), we get g(x) as described by Eq. (27).

The g(x) curve as depicted in Fig. 8 present a deceasing behavior then an increasing behavior. This phenomena is explained by the following mathematical analysis equations:

For  $g(x) = (-x + 1)^2 \exp(-x + 1)^2$ 

$$\lim_{x \to -\infty} g(x) = \lim_{x \to +\infty} g(x) = +\infty$$
(28)

$$\lim_{x \to 1} g(x) = 1 \tag{29}$$

Equations (28) and (29) shows the symetric behavior of g(x) where the curve has a minimum at the point A(1, 1).

Indeed, VSMN has a deceasing behavior from] –  $\infty$ , 1 [, then an increasing behavior from] 1,+  $\infty$  [.

Function g(x) is illustrated by the following equation, and the VSMN behavior is depicted in Fig. 8.

$$Y5 = (-Y4 + 0.5)^2 exp(-Y4 + 0.5)^2; p = q = 0.5, n = 2$$
(30)

#### 4th Case:

For the fourth layer, n = 3, g(x) is given by the following equations, and the SVMN behavior is provided in Fig. 9. Indeed, the output of each neuron will be the input of the next neuron for each layer, thus the cascade architecture of our model.

The VSMN curve has strictly a decreasing behavior, as depicted in Fig. 9(a) and (b). This behavior is explained by the following equations:



Fig. 8 VSMN behavior, n = 2: (a) VSMN curve, (b) Output USCT image



Fig. 9 VSMN behavior: (a) n = 3, p = q = 0.5, (b) n = 3, p = q = 0.75

$$g(x) = k(-x+q)^{n} exp(-x+p)^{2}$$
  
For n = 3, p = q = 1, k = 1, g(x) = (-x+1)^{3} exp(-x+1)^{2} (31)

For 
$$g(x) = (-x+1)^3 \exp(-x+1)^2$$
  

$$\lim_{x \to -\infty} \exp(-x+1)^2 = +\infty$$
(32)

$$\lim_{x \to -\infty} \left( -x + 1 \right)^3 = +\infty \tag{33}$$

From Eqs. (32) and (33) we get the Eq. (34) as follows:

$$\lim_{x \to -\infty} g(x) = +\infty \tag{34}$$

The g(x) function is illustrated by Eq. (35) when  $x \rightarrow 1$ 

$$\lim_{x \to 1} g(x) = 0 \tag{35}$$

$$\lim_{x \to +\infty} g(x) = -\infty \tag{36}$$

Through the fourth layer, g(x) is illustrated by the following equation for various conditions as follows: n = p = 0.5 and n = p = 0.75.

$$Y6 = (-Y5 + 0.5)^3 exp(-Y5 + 0.5)^2; p = q = 0.5, n = 3$$
 (37)

$$Y7 = (-Y6 + 0.75)^3 \exp(-Y6 + 0.75)^2; p = q = 0.75, n = 3$$
 (38)

🖄 Springer

$$Y8 = (-Y7 + 0.5)^{3} exp(-Y7 + 0.5)^{2}; p = q = 0.5, n = 3$$
(39)

As shown in the curves in, Figs. 5, 7, 8 and 9, the optimized VSMN presents a symetric behavior decreasing and then increasing for case 1 and case 3. Moreover, it shows a strictly decreasing behavior in case 2 and case 4, providing its efficiency to be applied on the USCT images. Consequently, the SVMN achieves its success to be applied in medical imaging area. To conclude, the deeper the neural SVMN, the higher the quality and the resolution of USCT images. Indeed parameter n has a great impact on the USCT image quality. When the number of n increases, the quality of images becomes better.'

### 3.3.1 VGG –SegNet model

Principle of proposed VGG-SegNet

VGG-SegNet is classified as a neural network for semantic segmentation. It is optimized in this work to segment the USCT images of bones. It was performed with 10 labels in [1]. In our work, we use four labels to segment different anatomical structures: the first for the background, the second for the cortical bone, the third for the cancellous bone and the fourth for the medullary cavity. It consists of two blocks: One plays the role of a coder and the other of a decoder. Each coder is made up of several layers. Its principle is illustrated by the application of the convolution accompanied by batch normalization followed with Relu activation layers. Then, passing through a pixel wise classifier layer and subsequently a soft-max layer. For the decoder block, it also consists of four resampling layers, a soft-max layer and 13 convolution layers with batch normalization and Relu, as depicted in Fig. 10 and detailed in Table 3. Accordingly, the sizes of convolutional kernels are set to 3\*3 for each five layers constituting the encoder and decoder blocks. These kernels perform a convolutional operation resulting in the output representing the map shape structure to detect in an input image. After each convolutional layer, an activation layer is added to perform a non-linear propriety, increasing the robustness of our VGG-SegNet model architecture. At the end, it will introduce maxpooling, which will detect the presence of characteristic cards in a region, hence the storage of each index containing the value extracted by each window. During the max-pooling phase the indices will be stored. It is a pre-learning phase. The encoder reduces the spatial dimensions thanks to the pooling layers, while the decoder reproduces the details of the image and the spatial dimensions. For the decoder block, it uses resampling, convolutions and the soft-max classifier. Resampling is performed on the inputs based on the indices stored during the encoding phase. Its principle is shown in Fig. 10. The result obtained at the decoder output will be transmitted to a soft-max classifier, which gives the final prediction, that will be an nchannel image.

### 3.3.2 VGG-U-net model

As compared to SegNet, the proposed U-Net does not reuse pooling indices but it rather transfers the entire feature map to the corresponding decoders and concatenates them to up sample via the decoder the feature maps via the decoder. There is no conv5 and max-pool 5 block in the U-Net.



Fig. 10 Internal architecture of VGG-SegNet

# 4 Results

### 4.1 VSMN implementation results

The VSMN implementation yields noise removal from USCT images, as shown in Fig. 11. As a consequence, it enhances the Signal to Noise Ratio (SNR) values, as provided in Table. 4.

Encoder			Decoder		
Block	Image Size	Filter	Block	Image Size	Filter
Block 1			Block 1		
Conv1+Relu	Input (256*256)	(64, (3,3)) (64,(3,3))	Up sampling	Input (14*14)	512,((3,3))
Conv2+Relu	Output (224*224)	2D (2,2)	Zero Padding	Output14*14	
Maxpooling			B.N		
Block 2			Block 2		
Conv1+Relu	Input (224*224)	(128, (3,3))	Up sampling	Input (28*28)	512,((3,3))
Conv2+Relu	Output 112*112	(128, (3,3))	Zero Padding	Output (28*28)	
Maxpooling		2D (2,2)	B.N		
Block 3			Block 3		
Conv1+Relu	Input 112*112	256,((3,3))	Up sampling	Input (56*56)	256,((3,3))
Conv2+Relu	Output 56*56	256,((3,3))	Zero Padding	Output (56*56	
Conv3+Relu		256,((3,3))	Conv2DLayer		
Maxpooling		2D (2,2)	B.N		
Block 4			Block 4		
Conv1+Relu	Input 56*56	512,((3,3))	Up sampling	Input (112*112)	128,((3,3))
Conv2+Relu	Output 28*28	512,((3,3))	Zero Padding	Output(112*112)	
Conv3+Relu		512,((3,3))	Con2D Layer		
Maxpooling		2D (2,2)	B.N		
Block 5			Block 5		
Conv1+Relu	Input 28*28	512,((3,3))	Up sampling	Input (224*224	64,((3,3))
Conv2+Relu	Output 14*14	512,((3,3))	Zero Padding	Output(224*224)	
Conv3+Relu		512,((3,3))	Con2D Layer		
Maxpooling		2D (2,2)	B.N		

Table 3	VGG-SegNet	architecture
---------	------------	--------------



Fig. 11 Results of VSMN implementation: (a): Layer 4 (b): Layer 5(c): Layer 6 (d): Layer 7 (For  $g(x) = (-x + 1)^3 \exp(-x + 1)^2$ ): Output

Furthermore, the VSMN increases the USCT image database, passing from 50 original USCT images to 400 augmented USCT images. As presented in Fig. 12, our approach allows us to present a free database for USCT researchers given the unavailability of these images and the difficulty to obtain them [28].

As depicted in Table 4, we present the mean SNR values of our USCT dataset used for the training, validation and testing processes. The yielded mean SNR values are illustrated as depicted in Table 4. The testing images shows SNR scores less than which are used for training and validation phases.

### 4.2 VGG-SegNet implementation results

#### 4.2.1 Dataset labeling

To automatically segment USCT data images, we have to annotate them using the Labelme tool for USCT image labeling with the Linux Operating System (OS). We label 400 USCT images. These annotations will represent the ground truth. Then 50% of images will be used for training, 25% for validation and 25% for testing. In fact, each image is segmented manually by a specialist radiologist into four regions. The first presents the background, the second shows us the cortical bone, the third presents the cancellous bone, and the fourth represents the medullary cavity. Figure 13 demonstrates an example of a manually labeled image.

### 4.2.2 Accuracy and loss results during the training and validation processes

Using the framework with the Linux OS, Python language, Keras and TensorFlow libraries with an Nvidia Titan X GPU, we train 200 USCT images with a size of (256\*256) with 10 epochs and 512 iterations for each epoch, from our USCT dataset on both the CPU and the GPU. The accuracy training results achieve 97.38% on the GPU and 89% on the CPU, but the validation results achieve 96% of accuracy, as shown in the screenshot of Appendix 1 and

Images	Number of images	Mean SNR
USCT images used for training	200	15.87
USCT images used for validation	100	15.42
USCT images used for testing	100	14.36

Table 4 SNR results of subsamples of USCT images





Appendix 2. Therefore, our code implementation on the GPU improves the accuracy results with a value of 8.38% compared to that implemented on the CPU as depicted in Table 5.

### 4.2.3 Models accuracy and loss curves during the training and the validation process

The loss and accuracy curves are important to determine the model behavior through training epochs, as it gives the direction in which the networks learn. The two curves as presented in



Fig. 13 USCT image labeling, (a): USCT bone image, (b): Annotated USCT, (c): USCT image mask

Fig. 14, using Adam optimizer network, show an excellent accuracy for both training and validation phases through 10 epochs and 512 iterations per epoch.

The two curves as depicted in Fig. 15, using Adam optimizer network, for both validation and training processes demonstrate a good fit as they represent a small gap between the two final loss values. The excellent achieved fitting is explained by the huge role that play the Adam optimizer network in decreasing the loss function and the process of USCT dataset augmentation, that show to be efficient to ovoid the over fitting to the training dataset. Moreover, The training loss curve decreases to a stability point. Furthermore, it shows to be lower on the training than on the validation, but the gap is too small. The validation loss curve has a small gap with the training.

### 4.2.4 Segmentation results

After having trained the USCT bone images, we have to automatically test the USCT images not seen before by the system. We use 100 images for validation and 100 USCT bone images for testing. The used dataset for the validation process is depicted in Fig. 16. The segmentation results of the USCT images used for validation achieve 96% of accuracy on the GPU and a high resolution of segmented images, as presented in Fig. 17. In fact, each USCT image represents three regions of interest showing the internal structure of bones, such as the cancellous bone and the medullary cavity, and the external bone structure, like the cortical bone, which represents the brown colour in Fig. 17. The comparison of the segmented images

Epochs	Train accuracy	Train loss	Validation accuracy	Validation loss
1	82.64%	0.519	82.66%	0.4764
2	83.43%	0.4488	83.56%	0.4603
3	84.10%	0.4181	79.86%	0.506
4	84.95%	0.3927	83.91%	0.393
5	87.28%	0.3203	86.11%	0.2008
6	92.67%	0.1919	89.95%	0.1474
7	95.33%	0.1288	94.17%	0.1274
8	96.43%	0.1014	95.12%	0.1215
9	96.99%	0.0888	94.53%	0.1450
10	97.38%	0.079	95.82%	0.1115

Table 5 Accuracy results during training and validation processes

13553



Fig. 14 Model accuracy during the Training and validation processes

used for validation with the ground truth keeps a high similarity between both types of images, as depicted in Fig. 18. For the USCT images used for testing, as illustrated in Fig. 19 and for Fig. 20, three regions are represented and the noisy background is removed. These tested segmented images show their efficiency by presenting a small error of 0.0061 compared with the ground truth and a high value of PSNR as detailed in Table 6, where the mean PSNR score is 10.44. Moreover, the segmented images are validated by a specialist who has ensured these results. The proposed method is validated by the following section in the discussion.

These results are yielded thanks to our proposed model architecture, which combines the neural VSMN network and the neural VGG-SegNet network. Indeed, the deep-learning-based neural network for automatic segmentation needs big data for images to achieve high accuracy of segmented images not seen before. Accordingly, the VSMN with its high architecture consisting of seven layers and four neurons for each layer automatically removes noise from USCT images.

#### 4.2.5 MSE, PSNR and IOU results



• PSNR

Fig. 15 Model loss during the training and validation processes



Fig. 16 Dataset of USCT bone images for validation

The PSNR shows its significance in determining the image quality reconstructed pixel by pixel. It is determined by the following equation.

$$PSNR = 10\log \frac{MAXI i^2}{MSE}$$
(40)

where MAXI i presents the maximum value in USCT images.

• MSE



Fig. 17 Segmented validation results



Fig. 18 Comparison of Segmented validation results with the ground truth, (a): Input USCT images, (b): Ground Truth, (c): Segmented USCT images

The Mean Square Error (MSE) makes it possible to determine the error existing between an original image and a reconstructed or segmented image [13, 36]. As depicted in Table 6, we obtain promising results.

• IOU



Fig. 19 USCT bone images used for testing



Fig. 20 Segmented USCT bone images used for testing

The IOU score is a standard performance measure for the segmentation problem. Thus, the IOU measures the similarity between the predicted segmented region and the ground-truth region for a set of images. It is defined by the following equation.

$$IOU = \frac{\text{area of overlop}}{\text{area of union}}$$
(41)

The process to segment USCT images of bones, using an end-to-end neural-network architecture, shows its efficiency in automatically determining the different anatomic bone structures with a high resolution. This contribution aims to facilitate the diagnosis process for clinicians, given the issue of analyzing the original noisy USCT images.

### 4.3 Implementation results of proposed model on GPU

Our framework is based on the Python language of the Keras package and on a Nvidia Titan X GPU using the Linux operating system. Graphics cards (GPU) are characterized by the large number of cores allowed by processors and the very large memory integrated with these processors. They are very useful for several computer tasks, precisely for software implementations like deep learning algorithms. Despite their energy consumption, GPUs show their efficiency given the success achieved in recent years for the implementation of deep

Parameters USCT used for the prediction process
Mean PSNR 10.44
Mean MSE 0.0061
IOU 0.96

 Table 6
 PSNR, MSE and IOU for subsamples of USCT bone images used for test

learning algorithms. As depicted in Table 7, VGG-SegNet requires weak memory for training and testing. The implementation of deep learning algorithms on GPUs is three times faster than their implementation on CPUs. The short time process implementation on GPUs is explained by the GPU architecture, designed for parallel graphics operations. Accordingly, the CPU and GPU architectures differ from each other. The CPU consists of multiple arithmetic and logic units, cache memory and dynamic random access memory. While the GPU consists of hundreds of ALUs, numerous control units, varied cache memory and DRAM memory [16, 33].

VGG-SegNet and VGG-Unet have the same memory inference and temporal process given the architecture used for both.

## 4.4 Proposed model evaluation

Our proposed Model is implemented on real scene images on both the CPU and the GPU, and the achieved results prove the robustness of our model which can be implemented on any dataset. Our suggested method keeps its efficiency to be applied to any database of images implemented on the basis of data from a real image scene and it shows good precision with very high robustness. The basis of the test images is presented in Fig. 21, which will be increased and segmented by our proposed neural network method. The obtained results demonstrate the robustness of our method, which can be implemented on any database (Figs. 21 and 22).

# **5** Discussions

The results show that physicians without a coding experience can use automated deep learning to develop algorithms that can perform clinical classification tasks at a level comparable to traditional deep learning models that have been applied in the existing literature. To validate the performance of our results, we determine the PSNR, MSE ratios and IOU score, as given in Table 6. The results of the MSE in the validation images show very small error values. Furthermore, for the PSNR, the values are encouraging due to the original image quality, so we can say that the PSNR is very improved for segmented ultrasonic tomographic test images. When comparing our segmented image results with the state of the art [14, 20], we succeed in solving the segmentation problem of ultrasound tomographic images with deep-learning, which offer us a free database. As depicted in Table 8, a comparative study with the state of the art is done with different neural networks applied on MRI, CT, and X-ray images of bones, given the unavailability of deep-learning work applied on USCT bones images. Moreover, the

Network	GPU training memory	GPU inference memory	GPU runtime in training	GPU runtime in testing	CPU runtime	Energy consummation
VGG-SegNet	10 MB	12,194 MB	1 s/step	0.15 s/step (Appendix3)	3 s/step	17 W/200 W
VGG-Unet	10 MB	12,194 MB	+	0.15 s/step (Appendix3)	_	-

 Table 7 Implementation results on GPU and CPU



Fig. 21 Real scene images

USCT dataset presents a big challenge [28], which prohibits the comparison of deep learning work applied on USCT images with ours. Our proposed model, by combining the optimized VSMN with VGG-SegNet, achieves 97.38% accuracy for the training phase and 96% for the validation phase and an error of 0.006 for the segmented test images. In fact, these results surpass those of the state of the art in [11], where the error exceeds 14% for the training phase



Fig. 22 Segmentation results of real scene images during the testing process on GPU

Neural network model	Train accuracy	Validation accuracy	Dataset		
SVMN-VGG-SegNet	97.38%	96%	Bone USCT images		
SVMN-VGG-Unet	97.38%	96%	Bone USCT images		
SegNet [25]	91.47%	_	MRI brain images		
CNN [26]	92%	_	CT bone images		
CNN-UNet	92%	_	CT scans bone images		
CNN [11]	85%	_	MRI vertebral bone		
Fully-Automated deep learning based CNN [22]	94%(1 year) 90%(1 year)	-	Human bones		
SegNet [29]	-	95%	CT lung images		
UNet [29]	_	91%	CT lung images		
VGG-SegNet [27]	95.86%	_	Lung CT Parenchyma images		
SegNet [5]	63.89%	_	Gastric cancer images		

Table 8	Accuracy	comparative	study with	state of	the art

and 20% for the validation phase during the segmentation process of MRI vertebral disc images. Accordingly, our proposed neural network overcomes the CNN [25] and SegNet [17] by a value of 6% due to our optimized architecture, as detailed in section 3. Moreover, our validation results are very promising compared to that was found by the state of the art in [29]. In addition, our VGG-Segnet proves to be excellent compared to which is implemented in [5] on gastric cancer images. Furthermore, our suggested method has reasonable accuracy with a small medical dataset.

# 6 Conclusion

This work presents an end-to-end neural network architecture called VSMN-VGG-SegNet, for the automatic segmentation of bones in USCT images for a short time process by a software GPU code implementation. The VSMN has proven its efficiency with an improvement of image resolution, a PSNR enhancment and noise removal. Moreover, it has performed free data for USCT researchers. Furthermore, the VGG-SegNet has provided excellent segmentation with an error of 0.006 applied on USCT images not seen before by the system. The robustness of our suggested model has demonstrated its robustness by achieving promoting segmentation results. Finally, the evaluation of our results has shown the efficiency of the proposed method in comparison with previous work. The next step will be dedicated for to the structure detection of USCT bone images for an automatic diagnosis using a deep-learning application on the GPU.

Funding This study has been funded by the ministry of high education and scientific research in Tunisia.

### Declarations

Conflict of interest Authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants performed by any of the authors.

Informed consent This article does not contain any patient data.

# References

- Badrinarayanan V, Kendall A, Cipolla R (2017) Segnet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell 39(12):2481–2495
- Belal SL, Sadik M, Kaboteh R, Enqvist O, Ulén J, Poulsen MH, Trägårdh E (2019) Deep learning for segmentation of 49 selected bones in CT scans: first step in automated PET/CT-based 3D quantification of skeletal metastases. Eur J Radiol 113:89–95
- 3. Bouallegue K (2017) A new class of neural networks and its application. Neurocomputing 249:28–47
- Bullock, J., Cuesta-Lazaro, C., & Quera-Bofarull, A. (2018) XNet: a convolutional neural network (CNN) implementation for medical X-ray image segmentation suitable for small datasets. arXiv preprint arXiv: 1812.00548.
- 5. Cai L, Gao J, and Zhao D (2020) A review of the application of deep learning in medical image classification and segmentation. "Ann Transl Med 8.11.
- Dahdouh S (2011) Filtrage, segmentation et suivi d'images échographiques applications cliniques, thèse université Paris Sud
- Dong J, Fu J, He Z (2019) A deep learning reconstruction framework for X-ray computed tomography with incomplete data. PLoS One 14(11):e0224426
- Drozdzal, M., Vorontsov, E., Chartrand, G., Kadoury, S., & Pal, C. (2016). The importance of skip connections in biomedical image segmentation. In deep learning and data labeling for medical applications (pp. 179–187). Springer, Cham.
- Fradi M, Youssef WE, Lasaygues P, Machhout M (2018) Improved USCT of paired bones using waveletbased image processing. International journal of image, graphics and signal processing, 10(9), 1.
- Fradi M, Youssef WE, Bouallegue G, Machhout M, Lasaygues P (2018) Automatic USCT image processing segmentation for osteoporosis detection. In international conference on the sciences of electronics, Technologies of Information and Telecommunications (pp. 372-381). Springer, Cham.
- Guerroumi N (2019) Segmentation automatique par apprentissage profond de la colonne vertébrale scoliotique à partir d'images de résonance magnétique (doctoral dissertation, École de technologie supérieure)
- Hopp T, Zapf M, Ruiter NV (2014) Segmentation of 3D ultrasound computer tomography reflection images using edge detection and surface fitting. In medical imaging 2014: ultrasonic imaging and tomography (Vol. 9040, p. 90401R). International Society for Optics and Photonics.
- 13. Hore A, Ziou D (2010) Image quality metrics: PSNR vs. SSIM. In 2010 20th international conference on pattern recognition (pp. 2366-2369). IEEE.
- Torres JSM et al. 2019 Linear filtering method for bone structures in computerized ultrasonic tomography images. 2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA). IEEE.
- Jue J, Jason H, Neelam T, Andreas R, Sean BL, Joseph DO, Harini V (2019) Integrating cross-modality hallucinated MRI with CT to aid mediastinal lung tumor segmentation. In international conference on medical image computing and computer-assisted intervention (pp. 221-229). Springer, Cham.
- Kayid A, Khaled Y, Elmahdy M (2018) Performance of cpus/gpus for deep learning workloads. The German University in Cairo
- Khagi B, Kwon GR (2018) Pixel-label-based segmentation of cross-sectional brain MRI using simplified SegNet architecture-based CNN. J Healthcare Eng 2018:2018
- Klein A, Warszawski J, Hillengaß J, Maier-Hein KH (2019) Automatic bone segmentation in whole-body CT images. Int J Comput Assist Radiol Surg 14(1):21–29
- 19. La Rosa F (2017) A deep learning approach to bone segmentation in CT scans (doctoral dissertation)
- Lasaygues P, Guillermin R, Metwally K, Fernandez S, Balasse L, Petit P, Baron C (2018) Contrast resolution enhancement of ultrasonic computed tomography using a wavelet-based method–preliminary results in bone imaging. In proceedings of the Int. workshop on medical ultrasound tomography (pp. 291-302).
- LeCun Y, Bengio Y, Hinton G (2015) Deep learning nature 521 (7553), 436–444 Google Scholar Google Scholar Cross Ref
- Lee H, Tajmir S, Lee J, Zissen M, Yeshiwas BA, Alkasab TK, Do S (2017) Fully automated deep learning system for bone age assessment. J Digit Imaging 30(4):427–441
- Lundervold AS, Lundervold A (2019) An overview of deep learning in medical imaging focusing on MRI. Z Med Phys 29(2):102–127

- Marwa F, Youssef WE, Machhout M, Petit P, Baron C, Guillermin R, Lasaygues P (2019) Automatic recognition processing in ultrasound computed tomography of bone. In medical imaging 2019: ultrasonic imaging and tomography (Vol. 10955, p. 1095514). International Society for Optics and Photonics.
- Minnema J, van Eijnatten M, Kouw W, Diblen F, Mendrik A, Wolff J (2018) CT image segmentation of bone for medical additive manufacturing using a convolutional neural network. Comput Biol Med 103:130– 139
- M. Oda, N. Shimizu, K. Kawakawa, Y. Nimura, T. Kitasaka, K. Misawa, M. Fujiwara, D. Rueckert, and K. Mori(2016) "Regression forest-based atlas localization and direction specific atlas generation for pancreas segmentation, «in International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 556–563, Springer.
- Omar A (2019) Lung CT parenchyma segmentation using VGG-16 based SegNet model. In *IJCA* (Vol. 178, no. 44, pp. 10-13).
- Ruiter NV, Zapf M, Hopp T, Gemmeke H, van Dongen KW (2017) USCT data challenge. In medical imaging 2017: ultrasonic imaging and tomography (Vol. 10139, p. 101391N). International Society for Optics and Photonics.
- Saood A, Hatem I (2021) COVID-19 lung CT image segmentation using deep learning methods: U-net versus SegNet. BMC Med Imaging 21(1):1–10
- Sedik A, Iliyasu AM, El-Rahiem A, Abdel Samea ME, Abdel-Raheem A, Hammad M, ... El-Latif AAA (2020) Deploying machine and deep learning models for efficient data-augmented detection of COVID-19 infections. Viruses 12(7):769
- Sudha S, Jayanthi KB, Rajasekaran C, Sunder T (2019) Segmentation of RoI in medical images using CNN-A comparative study. In TENCON 2019-2019 IEEE region 10 conference (TENCON) (pp. 767-771). IEEE.
- Tong T, Wolz R, Wang Z, Gao Q, Misawa K, Fujiwara M, Mori K, Hajnal JV, Rueckert D (2015) Discriminative dictionary learning for abdominal multi-organ segmentation. Med Image Anal 23(1):92–104
- 33. Tung DL, Sekiyama T, Negishi Y (n.d.) "Involving CPUs into Multi-GPU Deep Learning
- 34. Van Sloun RJ, Cohen R, Eldar YC (2019) Deep learning in ultrasound imaging. Proc IEEE 108(1):11-29
- Wang X (2016) Deep learning in object recognition, detection, and segmentation. Found Trends Signal Process 8(4):217–382
- Yalman Y, Erturk İ (2013) A new color image quality measure based on YUV transformation and PSNR for human vision system. Turkish J Electrical Eng Comput Sci 21(2):603–612

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