



# A computer-aided diagnostic system for liver tumor detection using modified U-Net architecture

Anum Kalsoom<sup>1</sup> · Muazzam Maqsood<sup>1</sup> · Sadaf Yasmin<sup>1</sup> · Maryam Bukhari<sup>1</sup> · Zian Shin<sup>2</sup> · Seungmin Rho<sup>3</sup>

Accepted: 21 December 2021 / Published online: 19 January 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

A multimedia-based medical decision-making system is an ultimate requirement in the medical imaging domain. In the healthcare sector, achieving quick and efficient results is one of the primary needs as well as a necessity for both doctors and patients. Liver cancer is increasingly spreading disease across the globe, and its timely diagnosis requires an accurate and precise tumor segmentation. However, it is very challenging to segment the tumors because of the variability in appearance, hazy borders, diverse densities, and sizes of tumors. Currently, deep learning-based approaches are being applied in a variety of domains which results in better performances. In this paper, a deep learning method based on modified U-Net referred to as Residual-Atrous U-Net (RA-Net) is proposed to segment the liver tumors. The suggested model is implemented by employing U-Net as a base model and extracts the tumor's features utilizing a parallel structure-based Atrous convolution block embedded in the original U-Net. This addresses the heterogeneity of tumors in terms of sizes as well as retaining the wider context without introducing any parameters to the model due to the various scales dilated kernels. Moreover, a strong, smooth, and non-monoatomic Mish activation function is deployed in this parallel Atrous block to bring nonlinearity. Besides this, the features of tumors are also extracted from Res Block simultaneously which learns the residual between the input and output feature maps via a skip link that jumps directly from input to output. This identity mapping in this block overcomes network degradation and increases performance. Later on, all these extracted features are fused to be sent to the network and hence improve the feature learning of the original U-Net. Furthermore, we have used the strategy of segmenting the tumors directly from the CT-scan instead of the two-stage process used by the majority of existing methods. We have performed experimental analysis over the 3DIRCADb dataset and according to the results shown by the proposed RA-Net, the Jaccard score which is a performance indicator stands out at 72%.

---

This paper is an extended version of our paper published in the Proceedings of the 2020 International Conference on Artificial Intelligence (ICAI), Las Vegas, USA, 27–30 July 2020.

---

Extended author information available on the last page of the article

**Keywords** Liver tumor detection · U-Net · RA-Net · Tumor segmentation · Parallel structure Atrous convolutions · Mish activation · Feature fusion

## 1 Introduction

Liver cancer is listed as a widely spread cancerous disease with the sixth-highest death rate. According to statistics, by World Health Organization, 0.16% of liver cancer patients faced death in Ethiopia [1]. A major increase in mortality rate has been observed in the past couple of years due to liver cancer, and hence it must be controlled by scientific contributions. Hepatocellular carcinoma or HCC is primary cancer that happens to be a root cause of 80% of deaths [2]. This percentage equals almost 700,000 people every year. Primary liver cancers are generally caused by cirrhosis. Alcohol consumption, hepatitis B and C viruses, or obesity leading to liver disorder serve as the main ground for cirrhosis. There are several imaging tests like ultrasound, MRI, and CT which help to diagnose cirrhosis but CT is the most common one [3]. CT or Computed Tomography gives comprehensive cross-sectional images of the abdomen which makes it inclusive of all tests. In these CT scan images, the homogeneity among tumor lesions and other surrounding lesions is a challenging problem for accurate segmentation of tumors [4]. This contributes to major inaccuracy in tumor segmentation. Therefore, it is suggested to make use of enhanced images for better results [5]. In a CT scan, the tumor is identified with a difference in pixel intensity of a particular area inside the liver. If this difference seems to be dark in the shade when compared to neighboring tissues inside the liver, it is referred to as hypodense [6]. In other cases, this difference is called hyper-dense when it turns out to be bright as compared to the surrounding liver. Moreover, it is a common observation that all conventional techniques which belong to manual tumor segmentation take a lot of time for processing [7]. This is not an ideal case when it comes to a clinical setup. The human liver stretches over the maximum limit of 150 slices in CT volume normally. To add more, it is found that there exists a lot of irregularity in the shapes of the lesion as well as low-level contrast in tumor and surrounding regions. Furthermore, the fluctuation in liver shape and size varies in patients [8]. In addition, currently in some recent approaches researchers are using Contrast-enhanced magnetic resonance imaging (CEMRI) which is another modality used to diagnose tumors especially benign tumors [9]. Keeping an account of all these factors, researchers all over the world are focusing on Computer-Aided Diagnostic (CAD) systems for the segmentation of liver tumors out of abdominal CT scan images [10]. Presently, all manual or semi-automatic methodologies to segment out tumors are not very robust and efficient due to their dependency on edge detectors [11]. Numerous deep learning-based algorithms have been introduced until now for automatic tumor segmentation, and one of such algorithms is Convolutional Neural Networks (CNN). CNN continues to be one of the most widely used approaches that have gathered the attention of many researchers in liver tumor segmentation [12].

Fully convolutional neural networks are an advanced version of CNN architecture that are utilized for semantic segmentation. Semantic segmentation works over the

mechanism of allotting specific labels to all pixels of an image [13]. CNN's architectures for semantic segmentation and all other deep CNNs architectures for image classification usually need to have semantic representations from an input image and such systems conventionally obtain rich contextual data by boosting convolution layers [14]. Convolution kernel requires particular contextual information through receptive regions [15]. Very tiny receptive fields stay connected to small or large scale features, in that order. However, convolution features fully focus on ROI for specific receptive fields and do not pay attention to the details of contextual data [16]. Years of research work over the topic of liver tumor segmentation has revealed that there happens to be thin lines and boundaries which surround lesions in CT or MRI [17]. This gives rise to various challenges which further add to inaccurate tumor detection and poor segmentation. Therefore, enough contextual information with improved feature learning regarding ROI must be worked on as a prerequisite. To deal with this scenario, many of the researchers have introduced patch-based methods [18]. These methods work by converting medical images into small patches and later on segmentation is performed. Similarly, numerous researchers suggested better deep learning models with enhanced feature learning to increase performance; nevertheless, certain gaps need to be filled to do so by improving the model's feature learning. There are various end-to-end segmentation models proposed by different researchers out of which U-Net is most commonly used but the other models such as SegNet, and their modified versions are also proposed to carry out the tumor segmentation [19]. Besides the segmentation models, some object-detectors-based approaches such as RetinaNet are also adopted for tumor detection especially for small tumors [20]. Further, liver tumor detection first requires accurate identification of the liver, and then the tumor is identified [21]. Even with the improvements made in CAD systems over a while, direct liver tumor segmentation continues to be an open challenge.

In this research, we are focusing on a multimedia-based medical diagnostic system that provides an end-to-end solution to help out physicians. We bring forward the RA-Net model which is an improved version of the U-Net framework. A lot of researchers have employed the U-Net framework [22] to perform segmentation but there is still a gap in its feature learning that is left unaddressed. In this paper, a U-Net with improved feature learning is designed to segment out the tumors from the CT-scan image. The proposed model incorporates the process of feature extraction of liver tumors from two different structures. The first structure or block employs parallel Atrous convolutions which process the input with multi-view kernels operate synchronously on the input to capture global contextual information of heterogeneous size tumors. These Atrous convolutions assist in acquiring the wider context without introducing extra parameters which are often increased in traditional convolutions due to large size kernels. This parallel structure of Atrous convolution simultaneously obtains a wider context of different size tumors, as well as Atrous convolutions of different scales/dilations operating on input, which is capable of extracting the information at different scales from the image. In short, the main aim of this parallel structure is to address the issue that occurs in the feature learning process of heterogeneous nature tumors because the performance greatly depends on the accurate feature extraction. Besides this, the nonlinearity is achieved by employing non-monoatomic and smooth activation namely Mish over the

resultant feature maps. Mish activation has very strong regularization effects and helps in controlling the overfitting of the model. This activation function is much better than ReLU because a ReLU completely ignores the negative values of feature maps. Further, for any node in the network whose activation due to ReLU is zero then all incoming and outgoing weights connected to that node are not updated. This is not happening with the Mish activation as it also takes the influence of negative values in feature maps. Moreover, the Res Block is also amalgamated which learns the residual mapping among feature maps resulting from convolution over the CT-slices using skip connections. The major rationale for incorporating this residual mapping is to prevent model weights from deteriorating. This generally happens when the number of layers in the model increases and hence the weight updates at initializing layers are approaching zero. In the end, the fusion operation aggregates all these features which are drawn from the analysis of CT-scan under different perspectives which include scales of context and residual mappings. Furthermore, most of those techniques worked on segmenting out the liver first, and then tumor segmentation is done. This is an easier approach as compared to tumor segmentation itself. However, our proposed method goes for tumor segmentation straightforward. It happens to be a challenging thing to achieve but the proposed RA-Net exhibits encouraging results. To determine the performance of the proposed model, we have tested the model on the publicly available 3Dircadb dataset. Moreover, we have also compared the proposed method with other existing methods. This research study contributes in the following ways:

- We propose an improved version of U-Net namely RA-Net to efficiently segment the tumors directly from the CT scan image
- To deal with heterogeneous shape and size tumors, RA-Net involves the parallel structure of Atrous convolutions in which multi-view dilated kernels assist in aggregating the wider context at different scales of CT-image
- Res-blocks in RA-Net are incorporated to obtain tumor features by learning residual mapping among feature maps stemming from convolution layers to address the problem of weight decaying during training
- The overall feature learning and performance of the original U-Net is further boosted up by gaining nonlinearity in the network using Mish activation having strong regularization effects.

The rest of the paper is organized as: Sect. 2 provides literature in the particular area. Section 3 discusses our proposed methodology along with details. Section 4 provides results and discussion. Section 5 of this research article provides the conclusion.

## 2 Related work

In this section, existing methods for tumor segmentation are discussed in detail. The existing methods in the context of tumor segmentation are categorized into different techniques as some researchers have utilized the traditional methods and some researchers employed deep learning-based methods. In traditional methods, Huang

et al. [23] employed the semi-automatic approach to segment the tumors. In their work, the feature representation of voxels is increased by applying the sampling of tumors in 3D space. Similarly, Wu et al. [24] also proposed the semi-automatic method based on improved fuzzy *C*-means (FCM) and graph-cuts approach to segment the tumors. They have also attempted to enhance the accuracy of segmentation by employing kernelized FCM with spatial information into the graph cuts methods and performing automated labeling of both foreground and background portions of images. Furthermore, Raj et al. [25] segments the tumor by first segmenting the liver from CT-scan using the Markov Random field method. This method provides noise resistance and rapid segmentation of tumors. Yang et al. [26] proposed the Kullback–LEIBLER divergence-based adaptive region growing and nonlinear mapping-based graph-cut method to segment liver tumors as a semi-automatic approach. Their proposed method is efficient in segmenting the tumors having poor contrast and weaker boundaries. There also exists some manual and semi-automated methods which indicates their reliance on various detectors e.g. edge detectors [27]. Some more examples of these techniques are the atlas-based models [28], deformable models [29], and graphical models [30]. All these semi-automated or traditional methods provide very accurate and precise segmentation of tumors but in real-time environments, they are not very much applicable due to the involvement of parametric steps. Furthermore, manual segmentation is not only dependent on human competence but also takes a long time. All these methods are now rarely used in automated applications. Fully automated methods would be extremely desirable in order to lessen the workload on medical operators.

Hence, in the context of the most advanced methods in tumor segmentation, deep learning models are the most widely used approaches in several research studies. Therefore, researchers' ultimate goal is automated tumor segmentation through deep learning methods. For instance, Christ et al. [31] segmented out liver and lesions using two cascaded deep neural networks and 3D conditional random fields. The methodology proposed in their work is comprised of two cascaded U-Net models for liver segmentation and tumor detection. The proposed framework requires the training of the cascaded model twice. More precisely, at the first stage, the liver is segmented and extracted ROIs are used to segment the tumors at a later stage. This process increases the overall time complexity of the algorithm. On the contrary, the dice similarity coefficient score attained by the proposed model is very encouraging and was nearly around 0.943 in liver segmentation. However, they have not incorporated tumor segmentation results in their research article. For this reason, we get no specific information about the results of tumor segmentation. Subsequently, Tran et al. [32] modified the standard version of U-Net by incorporating the dilated dense connections among the units of the networks. The skip connections are established using output responses from convolutional units. The proposed method exhibits the 73.34% value of dice score, however, it is stated that the connections among the nodes of the network are cumbersome if the count of convolutional units doubles. Similarly, Seo et al. [33] proposed another variant U-Net model namely, mU-Net to segment out the tumors and achieve a dice score of 68.14%. They have used the MSE loss function for the training of their model, however, the performance of the proposed model is further improved if some other loss function is used because as

stated in their work that MSE fails to capture the structural similarity. Later on, Liu et al. [34] designed the novel architecture namely GIU-Net in which graph-cut algorithm and U-Net are combined. They used a strategy of enhancing the depth of structure for improved semantic segmentation output. One more thing which is evident from their research study is that they introduced skip connections to be from the pooling layers output, and integrated it with a graph-cut technique. The evaluation measures used for the proposed method include DSC and its value was recorded to be 0.9505. Following that, Bai et al. [35] proposed another variant of U-Net architecture for accurate segmentation of liver tumors and achieved a stunning performance of 76.4% dice. In their proposed model, candidates are generated by the multi-scale candidate generation method (MCG). This MCG method is operated on segmented livers ROI from CT scan. In the end, a 3D fractal residual network (3D FRN) is used to detect the tumor areas followed by a post-processing step of the active contour method (ACM). It is used to refine the segmentation results of tumors. Furthermore, Budak et al. [36] also made use of a Convolutional Neural Network for liver tumor detection and segmentation. According to their research study, the proposed model was composed of two cascaded encoder-decoder architectures namely EDCNN. This model has two encoder and decoder modules with ten convolutional layers along with batch normalization and ReLu activation followed by a max-pooling layer. Moreover, one cascaded deep neural network is designed to segment the liver while the other one is used to segment the tumors from the segmented results of the liver. The resultant DSC values of liver and tumor segmentation obtained in this study were 0.9522 and 0.634, respectively.

In all of the above-mentioned existing approaches toward liver tumor segmentation have employed several methods such as traditional techniques, that give very good segmentation results, but these methods are not suitable for real-time scenarios due to the manual setting of several parameters. Following that, deep learning-based architectures are proposed by different researchers in which they attempt to enhance feature learning using different configurations among layers of architectures. However, the heterogeneity of tumors is not taken into account in all of these existing methods. Some segmentation methods have extremely strong performance, but the vast number of layers creates a high level of computational complexity and the problem of weight decaying. The feature learning process does not take into account a broader context and extracts features from tumors of varying sizes. Hence, a gap exists to improve the feature learning by maintaining model complexity lower in terms of parameters. Therefore, in continuation to these deep learning approaches, we seek to propose another novel deep learning algorithm namely RA-Net by modifying the feature learning process of the original U-Net [37] as our primary contribution. The specialties of the proposed algorithm make it effective to segment the different heterogeneous nature tumors such as different, shapes and sizes. Moreover, in most of the existing studies, tumors are segmented from liver ROI after the liver segmentation. Here in this research, we strive to segment the tumors directly from CT scans as well as make the algorithm a holistic end-to-end solution to segment the tumors. Furthermore, this research study is helpful for both, physicians as well as patients because it works well over both dimensions.

### 3 Proposed methodology

A pictorial overview of the proposed methodology is shown in Fig. 1. It is clear from Fig. 1 that the proposed framework comprises several steps. At first, the data of liver tumor patients in form of CT scans are acquired. Later on, the acquired CT scans are passed through preprocessing stage to enhance their visibility. Subsequently, these preprocessed CT-scan images are further given as an input to the proposed RA-Net which accurately segments the tumor and classifies each pixel in the CT-scan image into background class or tumor class. All these steps are explained in detail below.

#### 3.1 Dataset acquisition

In this research study, the 3Dircadb dataset is utilized to access the performance of the model. This dataset is also referred to as 3D image Reconstruction for Comparison of Algorithm Database. In this dataset, 20 folders are comprised the CT scans of tumors from several European hospitals. More specifically, it consists of a CT scan of the liver for 20 patients in which 75% of patients are diagnosed with hepatic tumors. The CT scan images are included in DICOM format in the dataset along with their associated label as binary images. Each 3D CT scan of the patient consists of a different number of CT slices. Moreover, there also exists some slices where the tumor is not present that is also taken into account during experimentation. In addition, the dimension of 2D CT slice images is  $256 \times 256 \times 1$  which are given as an input to the proposed RA-Net during training. The more specific details of the dataset are given in Table 1.

#### 3.2 Preprocessing of CT-scan images

Usually, the datasets of medical imaging contain noisy textures which ultimately caused the ROI to fade away. There are different kinds of noises which include blotches, uneven patches, unwelcome objects, and organs, etc. So, to prepare the dataset for experimentation, this data must be preprocessed. Further, this preprocessing step enhanced the raw images so that tumor features are highlighted. Different researchers have employed several methods in this step to enhance the visibility of medical images. In this research study, we have employed the contrast enhancement using windowed Hounsfield unit values within  $[-100, 400]$  range. This method results in clear CT

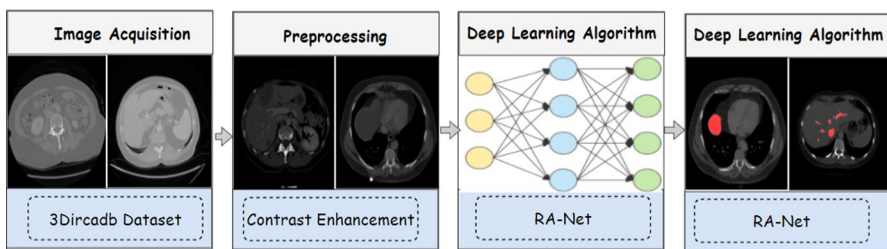


Fig. 1 Pictorial Overview of the proposed framework for liver tumor detection



**Table 1** Dataset details used in this research

S. no.	Gender	YOB	Liver pathologies	Liver size	Image size
1	M	1966	0 tumor in liver	19,1	512
				14,4	512
				16,2	132
2	F	1973	1 tumor in VI	19,3	512
				17,7	512
				11,0	260
3	M	1951	20 tumors	20,0	512
				12,9	512
				18,1	122
4	F	1970	0 tumor in liver	22,4	512
				15,4	512
				13,4	113
5	F	1946	2 tumors in II and VIII	18,8	512
				17,7	512
				14,7	125

images which are free from any type of noisy blotches, uneven patches, organs, and undesirable objects. Moreover, this method of preprocessing is also adopted by various researchers in this field [36, 38]. Some example images of CT scans before and after preprocessing are shown in Fig. 2.

### 3.3 Proposed Residual Atrous U-Net (RA-Net)

The proposed algorithm namely Residual Atrous U-Net (RA-Net) is designed to segment the tumors directly from the CT scan image of the patient. It mainly consists of three main parts, namely contracting path, bottleneck path, and Expansive path. The detailed explanation of each path is discussed in detail below and a pictorial representation of the proposed RA-Net is shown in Fig. 3. Furthermore, the symbols with their definition used in the following sub-sections are described in Table 2.

#### 3.3.1 Contracting path

The working of RA-Net starts from the contracting path by giving the input of CT slices of size  $256 \times 256 \times 1$ . This input is passed to the RA block of the proposed RA-Net as shown in Fig. 4. The RA block is comprised of Res block and parallel structure-based Atrous convolutions. After the RA-block, the feature maps are passed to the max-pooling operation of size (2, 2) to downscale an image. To control the overfitting dropout layer of the rate 0.05 is added after every max-pool operation. The output of the max-pool operation is defined by an Eq. (1) which is simply a maximum value from a given input signal.

$$y_{k,w}^i = \max_{0 \leq a,b \leq p} (x_{k \times p+a, w \times p+b}^i) \tag{1}$$



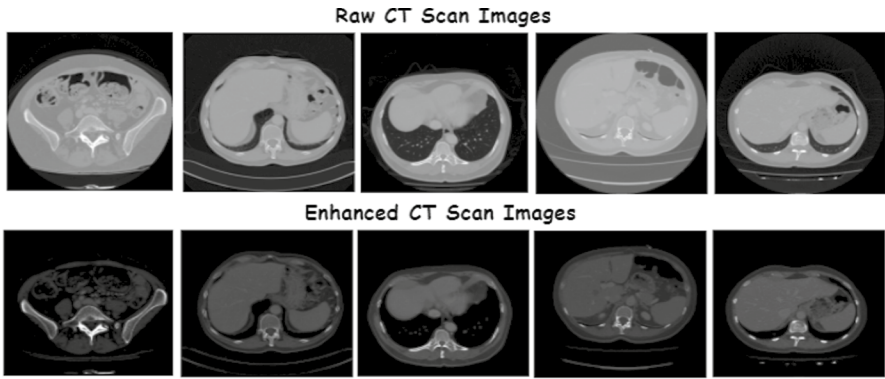


Fig. 2 Sample raw and enhanced CT-scan images

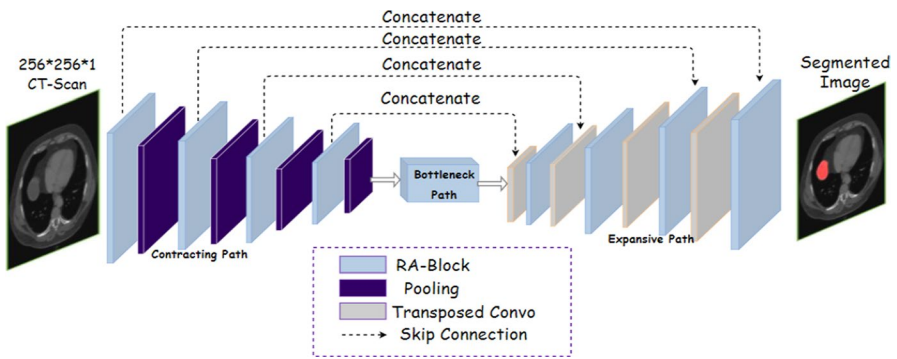


Fig. 3 Proposed Residual-Atrous U-Net (RA-Net) for liver tumor segmentation

Table 2 Notations and definitions

Notations	Definitions
$x$	Feature map resulted from convolution
$x_i$	Representing particular location $i$ on a feature map
$w_{i,j}$	Weight between node $i$ and $j$
$\sigma$	A term representing ReLu activation
$b_j$	Bias term for node $j$
$W_s$	Linear projection
$y_{k,w}^i$	A neuron $y$ present on a position $(k, w)$
$p \times p$	Size of pooling window during down-scaling
$r$	Denotes the stride size
$e^x$	Exponential function
$y$	Output feature map
$k$	Kernel size
$f'(x)$	The first derivative of function $x$

In Eq. (1), at the  $i$ th output map, a neuron  $y_{k,w}^i$  is present on position  $(k, w)$  at the down-sampling layer. The maximum value is assigned to a neuron  $y_{k,w}^i$  in the  $i$ th input map  $x_i$  from region  $p \times p$ .

### 3.3.2 Architecture of Residual Atrous (RA) block

As shown in Fig. 4, the input toward the RA block is passed further into two blocks i.e. parallel-structure-based Atrous convolution block and Res block synchronously. In parallel Atrous convolution block, every Atrous convolution of size  $1 \times 1$ ,  $2 \times 2$ , and  $3 \times 3$ , is used to capture the wider contextual information from multiple views. This block helps in grabbing the tumor’s information of different sizes and scales. For Atrous convolutions, consider a feature map  $x$  and weight matrix called kernel  $w$  and an input signal of 2D, then for every location  $i$  the  $y$  output is computed as follows:

$$y[i] = \sum_k x[i + r \cdot k]w[k] \tag{2}$$

In the above Eq. (2), the parameter  $r$  represents the strides through which an input signal is sampled and it is equal to convolve over the input  $x$  using kernels  $w$ . In between every consecutive filter value, there is the insertion of  $r - 1$  zeros through which the kernels are up-sampled along every spatial dimension. The Atrous convolutions help to enlarge the field of view of filters without the drop of information about loss and coverage as shown by Yu et al. [5]. Moreover, in comparison with traditional convolution, this Atrous convolution extracts the wider context with fewer parameters. E.g. a  $3 \times 3$  kernel is convolved over the input feature map of  $7 \times 7$ . In this case, a  $3 \times 3$  kernel is exhibiting the same effect as of  $5 \times 5$  to capture broader context but in this scenario, the total number of parameters used to achieve this step is 9 which is much lesser than 25 (in the case of  $5 \times 5$  kernel). Hence in the proposed parallel structure of Atrous convolution, the tumor features from the wider

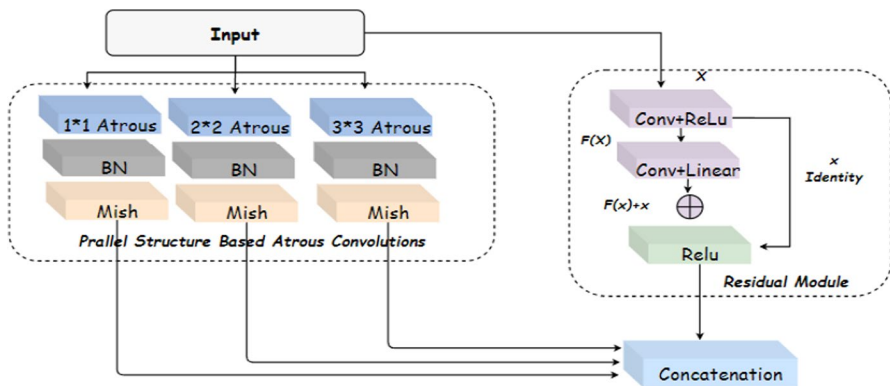


Fig. 4 The architecture of proposed RA-block

context of multiple views are extracted without incorporating any extra complexity in the network in terms of parameters.

After Atrous convolution is applied over the input, we have used the batch normalization and Mish activation function to bring the nonlinearity. This activation function is defined by an Eq. (3):

$$f(x) = x \tanh(\text{soft plus}(x)) = x \tanh(\ln(1 + e^x)) \tag{3}$$

The properties of the Mish activation function include that it is a continuous, smooth, non-monoatomic, and self-regularized activation function. The Mish activation function is with a range of  $[\approx -0.31, \infty)$ . In the case of bounded below and unbounded above, its derivative is defined in Eq. (4):

$$f'(x) = \frac{e^x \omega}{\delta^2} \tag{4}$$

where  $\omega = 4(x + 1) + 4e^{2x} + e^{3x} + e^x(4x + 6)$  and  $\delta = 2e^x + e^{2x} + 2$ . The sample graph of Mish activation in range  $(-10, 10)$  is shown in Fig. 5. Similarly, the second part of the RA block is the Residual module. As mentioned earlier, the input is also passed to the residual module. For every residual module used in our RA-Net, there are  $N$  elements at the specific or same position of the feature map, passes first through a convolution of size  $1 \times 1$ , so, we have  $N$  new elements at corresponding positions on feature maps as output. At a certain position of  $i$ th input channel we have an element value represented by  $x_i$ , and at the same corresponding position of  $j$ th output channel the output value is represented by  $x_j$ , and  $w_{i,j}$  is the weight between  $x_i$  and  $x_j$  then, and bias term is represented by  $b_j$  then the formula is:

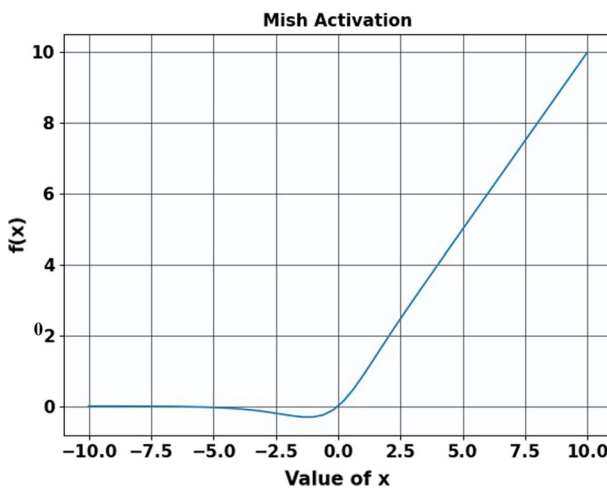


Fig. 5 Graph of Mish activation function in range  $(-10, 10)$

$$x_j = \left( \sum_i^N w_{i,j} * x_i \right) + b_j \quad (5)$$

This 1\*1 convolution is used to reduce the number of filters associated with the last layer and increases the filters at the input layer and hence it serves as a projection layer. This approach is utilized by the proposed RA-Net and it is called a projection shortcut. Therefore, a residual module in the RA block is defined as:

$$y = F(x, \{W_i\}) + x \quad (6)$$

In Eq. (6), the vectors of the input and output layer are represented by  $x$  and  $y$ . The term  $F(x, \{W_i\})$  represents the residual mapping that is to be learned. If we have two layers  $F = W_2\sigma(W_1x)$  in our residual module as shown in Fig. 4, in which the term  $\sigma$  represents ReLu activation function and an  $F + x$  operation is performed with the help of addition and shortcut connection. In Eq. (5), the shortcut connection neither incorporates any extra parameter and nor the computational complexity of the network. In Eq. (6),  $F$  and  $x$  are only added if their dimensions are equal. In case when dimensions are not equal then to match the dimensions, the shortcut connection performs a linear projection  $W_s$  and it is given by Eq. (7):

$$y = F(x, \{W_i\}) + W_s x \quad (7)$$

Subsequently, all these features that are drawn from either parallel structure-based Atrous convolution block or residual module are accumulated together by the concatenation operation. These fused features are further passed to the network to the succeeding layers. Moreover, in the contracting path, we have four RA-blocks followed by max-pool and dropout layers. Each RA-block is comprised of a parallel Atrous convolution block and residual modules. More specifically, the feature maps resulting from the first RA-block have dimensions of  $256 \times 256 \times 64$ , the dimensions of feature maps from the second RA-block is  $128 \times 128 \times 128$ , similarly, the feature maps drawn from the third and fourth RA-block are dimensions of  $64 \times 64 \times 256$  and  $32 \times 32 \times 512$ , respectively. A max-pool layer is added after every RA-block of window size (2, 2). The weights of all convolutional layers either in Atrous block or residual block are initialized with the ‘‘he\_normal’’ [39] weight initialization method. In addition, all convolutional layers are padded and stride size in both convolution and max-pooling layers is 1.

### 3.3.3 Bottleneck path

After the input is passed from the contracting path, the resulting feature maps are passed to the bottleneck path. In terms of deep learning, there is normally a layer comprised of very few nodes in comparison with prior layers. The bottleneck in a neural network is considered as a layer with a minimum number of neurons then layers before and after it. In deep learning, this layer is used to attain the representation of input with reduced dimensionality. Bottlenecks refer to forceful learning of

models about compression of inputs [40]. In the proposed RA-Net model, the bottleneck path consists of an RA-block which is comprised of parallel Atrous convolutions and residual module. The dimensions of feature maps resulting from RA-block in bottleneck path is  $16 \times 16 \times 1024$ . As indicated in the contracting path, the rest of the parameters such as stride size, weight initialization, and padding parameters remains the same. All these feature maps resulting from this path are passed as input to the upcoming expansive path.

### 3.3.4 Expansive path

The output returns from the bottleneck patch are provided as input to the first transposed convolution layer of the size  $3 \times 3$  followed by RA-block. The transposed convolution is padded and the size of the stride in this convolution is set to (2, 2). There are four transposed convolution layers and four RA blocks in the expansive path. After every transposed convolution layer, a skip connection is added which aggregates the outputs of the contracting path to the expansive path as shown in Fig. 3. This can be done by concatenating the output feature maps of transposed convolutions to corresponding feature maps of the contracting path. More specifically, the output of the first transpose convolution layer in the expansive is concatenated with the output of the last RA-block in the contracting path. The same procedure is followed for the reaming three transpose convolution layers. This concatenation operation helps in combining the localization information of the image with contextual information. The output dimensions of all RA-blocks in the expansive path are  $32 \times 32 \times 512$ ,  $64 \times 64 \times 256$ ,  $128 \times 128 \times 128$ , and  $256 \times 256 \times 64$ , respectively. At the last, the  $1 \times 1$  convolution layer is added with sigmoid activation to obtain a segmented image. More precisely, the reason for this  $1 \times 1$  convolution is to map the output of the last RA-block having the 64-component feature vector to the specified number of classes' i.e.tumor class and background class.

## 4 Experiments and discussions

In this section, the details regarding performance evaluation mechanisms and results of the proposed model are discussed along with a comparison to existing methods toward tumor segmentation.

### 4.1 Performance evaluation

In this research study, the following performance measures are used to access the performance of the model in segmenting tumors. These metrics include the dice similarity coefficient (DSC), Jaccard score, Symmetric Volume Difference (SVD), accuracy, and specificity.

#### 4.1.1 Dice similarity coefficient (DSC)

DSC is one of the most commonly used statistical parameters which measure the similarity of two sets of data [41]. In this research work, we have used DSC as a primary performance measure. Mathematically, it is defined as an area of overlap among two binary images. These binary images represent the actual ground truth of the CT-scan image and the predicted segmentation of the RA-Net. DSC is often reported in a range ranging from 0 (no overlap) to 1 (perfect overlap). Equation (8) shows the mathematical formula of DSC:

$$DSC = \frac{2TP}{2TP + FP + FN} \quad (8)$$

#### 4.1.2 Jaccard similarity score (JSC)

The Jaccard similarity score is another popular performance measure for segmentation algorithms. It measures the similarity between the actual ground truth of the CT-scan image and the predicted segmentation of the model. Its mathematical formulation is given in Eq. (9):

$$JSC = \frac{TP}{TP + FP + FN} \quad (9)$$

#### 4.1.3 Symmetric volume difference (SVD)

The SVD score is used to measure the discrepancy between actual ground truth images and the images segmented by the model. If SVD is equal to zero, it signifies a promising resulting segmentation value. The following Eq. (10) is used to compute the SVD score.

$$SVD = 1 - DSC \quad (10)$$

#### 4.1.4 Accuracy

Accuracy is one of the most widely used metrics to measure the overall performance of the model [42, 43]. The mathematical equation for computing the accuracy is given in Eq. (11):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

### 4.1.5 Specificity

Specificity is defined as the correctly recognized fraction of true negatives and is determined by an Eq. (12):

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (12)$$

In all of the above Eqs. (8) to (12), the TN represents the True Negative, TP represents the True Positive, FP represents the False Positive and FN represents the False Negative.

## 4.2 Results and discussion

In this research study, a deep learning model namely RA-Net is proposed to classify each pixel of CT-scan image into tumor or background class pixels. The performance of the proposed framework is assessed on the preprocessed CT scans of the 3DIRCADB dataset. Two non-overlapping sets of ratios 80–20 (train and test) are formed by randomly partitioning the data of 20 patients. The network's hyperparameters include the learning rate, which is set to 0.001, and the Adam optimizer, which is used to optimize the weights with a batch size of 4. The network simulations are run on 12 GB NVIDIA Tesla K80 GPU provided freely by Google Colab. Furthermore, we run the algorithm five times and then take an average to report the results. The results of the proposed model in terms of performance metrics that are described in the previous section are shown in Table 3. From Table 3, it is evident that the proposed model shows encouraging results in localizing the tumors. The average dice score attained by the model is 81% while the average Jaccard score is 72%, respectively.

Similarly, the values of other evaluation metrics are also computed. The SVD demonstrates the discrepancy between the actual and masks predicted by the model. As shown in Table 3, the suggested model obtained an accuracy of 93% with an SVD score of 0.19, which was the very minimal difference between the actual and images segmented by the model. The larger difference among accuracy and other metrics is due to the class imbalance. In the CT-scan slice, the total number of pixels that belong to the background class is larger than the total number of pixels that belong to tumor pixels. Hence, in the case of accuracy, the results might be biased toward the background class because accuracy measures the overall performance of all classes collectively. Therefore, the metrics that accurately describe the performance of RA-Net are Dice and Jaccard scores. Furthermore, the value of specificity

**Table 3** Results of proposed model in comparison with U-Net

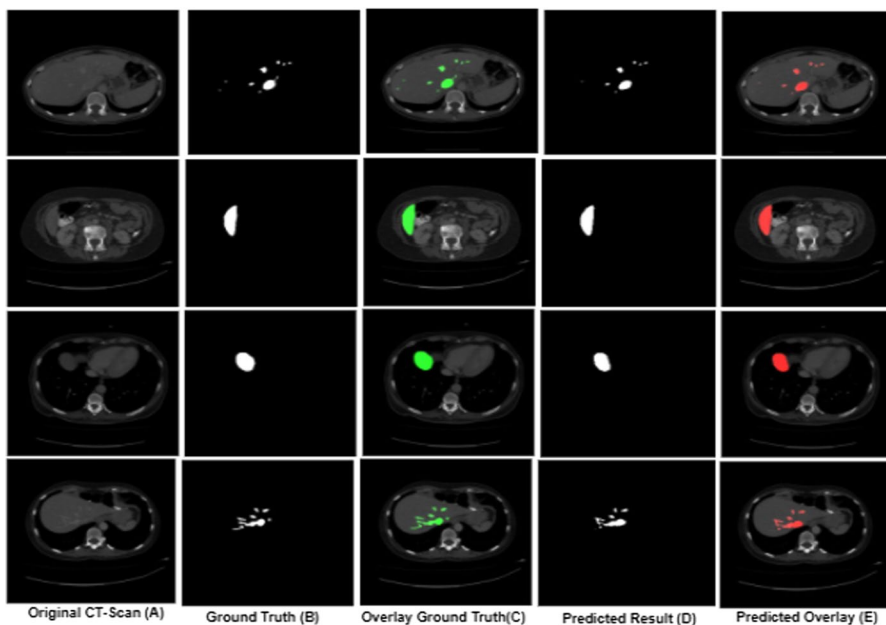
References	Dice score (%)	Jaccard (%)	Accuracy (%)	Specificity (%)	SVD
UNET [37]	67.5	56.0	92.6	70.1	0.33
Proposed method	81	72	93	82	0.19



is 82%. In addition, the performance of the proposed model is also compared with the original U-Net designed by Ronneberger et al. [37]. This standard architecture of U-Net is widely used by different research studies because it performs very well on tasks of biomedical image segmentation. In comparison with the proposed framework, the results of the standard U-Net for tumor segmentation are very low. More specifically, the dice and Jaccard score attained by the U-Net model is 67.5% and 56% on the same 3DIRCADB dataset. Similarly, the other scores attained by U-Net are specificity which is 70.1% while the SVD and accuracy are 0.33 and 92.6%, respectively. It is observed that the proposed model shows some improvement on standard U-Net and exhibits encouraging results.

Furthermore, some sample results of test CT-scan images are shown in Fig. 6. In Fig. 6, the test CT scan images of patients are shown in column (A), the actual ground truth images of CT scans are shown in column (B), the actual overlay results of ground truth over the CT-scan are shown in column (C), the CT-scan images that are segmented by the model are shown in column (D) while column (E) shows the segmented overlay results of CT-scan images. All the performance measures which include the SVD, DICE, Jaccard are computed by utilizing binary images present in columns (B) and (D) of Fig. 6.

Following that, we have also evaluated the performance of the proposed model and standard U-Net in terms of accuracy and loss during training. Accuracy is one of the ways which determines the efficiency of proposed models. Figure 7 (first graph) shows the accuracy of the proposed RA-Net and standard U-Net during training. The X-axis of the graph provides detail about several epochs and Y-axis shows



**Fig. 6** Tumor segmentation results of proposed RA-Net

accuracy values. Similarly, the model loss is also analyzed over a total number of epochs. Loss indicates how unsatisfactory a model's segmentation is over a single CT slice of a dataset. Figure 7 (second graph) shows the loss values over epochs of both the proposed model (RA-Net) and standard U-Net. In this Fig. 7 (second graph) the number of epochs is plotted over the  $x$ -axis, and values of loss are plotted on the  $y$ -axis.

### 4.3 Comparison with existing methods

After detailed experimentation, it can be observed that the proposed model performed very well and provided promising resultant values as well. However, it is important to perform a comparative analysis with some existing methods of tumor segmentation. This comparison fairly indicates that the proposed RA-Net is an efficient model that segments out liver tumors accurately. Table 4 gives an account of various researchers with resultant values of their proposed models for liver tumor segmentation. The resultant values show that the proposed modified version of U-Net exhibit encouraging results. The literature revealed that by utilizing the 3DIR-CADB dataset, a very good dice score of 61% is achieved in the work of Christ et al. [38] by using the two FCN model runs in a cascading manner. However, in their work, the tumor segmentation is accomplished after segmenting the liver from the whole CT-scan, hence the FCN first runs to segment the liver and to prepare liver ROI. Later on, the liver ROI is used as an input to again train the FCN to segment out the tumor. Therefore, the FCN model must be trained twice which increases the time of the algorithm. Similarly, a very encouraging dice score of 74.96% is attained by Alirr et al. [44] by employing adaptive thresholding-based methods to segment the tumors. Their proposed approach shows very remarkable improvement but due to its semi-automated nature user interaction is required to accurately segment the tumors. Subsequently, Bai et al. [35] have made some architectural modifications to the standard U-Net, and have attained the 76.4% results in terms of Dice score. Nevertheless, the approach of Bai et al. [35] requires the additional post-processing steps to refine the final results of tumor segmentation using the active contour

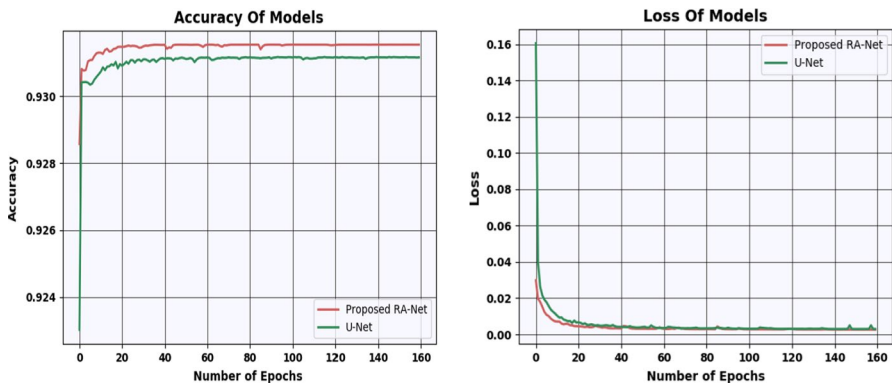


Fig. 7 Loss and accuracy graphs of standard U-Net and RA-Net

**Table 4** Comparison of the proposed model with state-of-the-art techniques

References	Dice score (%)	SVD	Complexity	Year
Budak et al. [36]	63.4	0.37	Two-stage training + 0.7 M Parameters + NVIDIA Quadro M6000 GPU + 24 GB memory, 3072 CUDA cores	2020
Christ et al. [38]	61	0.39	Two-stage training + NVIDIA Titan X GPU with 12 GB VRAM	2017
Alirr et al. [44]	74.96	0.26	Two-stage segmentation	2018
Tran et al. [32]	73.34	0.26	Two stage training + 21.7 M parameters + Intel Xeon Silver 4114 CPU, GRID Virtual GPU V100D-8Q, and 32 GB of RAM memory	2020
Seo et al. [33]	68.14	0.32	Intel Xeon E5-2698 v4 2.2 GHz (20-Core) CPU and two of four total Tesla V100 (32 GB memory for each GPU) GPUs	2019
Bai et al. [35]	76.4	0.24	Two stage training + NVIDIA GeForce GTX 1080ti GPU with 3584 cores and 11 GB memory	2019
Han et al. [45]	68	0.32	Two stage training + NVIDIA Tesla K80 graphic processing units (GPUs)	2021
Zhang et al. [46]	78.8	0.22	Two stage training + 5.3 M parameters	2021
Proposed method	81	0.19	Single Time training + 13 M parameters + 12 GB NVIDIA Tesla K80 GPU (Google Colab)	-

model (ACM). Subsequently, Budak et al. [36] have attained the dice score of 63.4% by using cascaded encoder-decoder architecture. Their method is also run in two stages to segment the tumors. Moreover, if the comparison is made with some recent works, Tran et al. [32] proposed the modified version of U-Net by incorporating the dense connections in the layers and achieves a very remarkable score of 73.34%. However, one of the major issues they discussed in their proposed model is that the model's connectivity becomes more sophisticated with the rise in the number of convolution units. Furthermore, Seo et al. [33] also proposed the variant of U-Net and have achieved the 68.14% dice score. Their proposed model uses the MSE-based loss function to train the model. They have stated that due to MSE loss, the model finds it difficult to capture the structural similarity of the tumors. However, if the MSE loss is replaced with another loss function, the performance may be increased even more. Besides this, some more latest studies on liver tumor segmentation include the work of Han et al. [45]. In their study, a boundary-loss-based fully-convolutional neural network is designed to segment the tumors with a dice score value of about 68%. Similarly, in the work of Zhang et al. [46] a Hybrid-3DResUNet is proposed to segment the tumors using 3D convolution operations. Their proposed model shows a very excellent performance in terms of dice score i.e. 78.58%. However, during inference time, the proposed model used a sliding window approach to generate 3D patches using volume data.

#### 4.4 Discussions

It is observed from the above results and comparison with existing methods that the proposed model shows encouraging results in tumor segmentation. As indicated before in the introductory section, the tumors have heterogeneous characteristics, i.e. irregular forms and sizes which degrades the effectiveness of the segmentation algorithm and segmenting such tumors is very challenging. Hence, to extract the features of different size tumors, we proposed the parallel structure-based Atrous convolution block. This block contains Atrous convolutions of different dilations/scales/sizes that best capture the information of different size tumors. This suggested block addresses the issue of encoding features of heterogeneous tumors by different scales kernels. Furthermore, we employed Atrous convolution rather than ordinary convolution to capture a broader context without incorporating extra parameters into the model because a  $5 \times 5$  kernel matrix in standard convolution can have 25 parameters. However, due to dilation rates, Atrous convolution acquires information from  $5 \times 5$  regions of the image with the  $3 \times 3$  kernel containing just 9 parameters. In this method, a larger view/context is targeted, which eventually increases the model's receptive field with a smaller number of parameters. Furthermore, another important aspect that plays a key role in the performance of the model is the activation function. In this proposed block, a Mish activation function is used instead of ReLu. The reason behind using Mish is its strong regularization ability. More precisely, the ReLu activation completely ignores the negative values in the features maps which may loss of information and the weights goes inward and outward from a neuron on which

the ReLU outputs the values zero are not updated. However, in comparison with ReLU, Mish does not ignore the negative values in feature maps, hence the information is preserved. In addition, if we improve the architecture from the perspective of layers configuration, then the larger number of layers causes the network to overfit and ultimately the decaying of weights occurs. Hence, the problem of network degradation is also addressed by learning residual mapping among feature maps stemming from convolution layers. It is logically concluded that we attempt to improve the U-Net in three different aspects, the first aspect is the challenge of accurate feature extraction from heterogeneous size tumors, the second and third aspect is the architectural level improvements, such as the problem of weight decaying, overfitting, and choosing essential information from features maps. All of these qualities make the model better to learn and segment the tumor more accurately which is the principal cause for improved performance. Besides this, the proposed model directly segments the tumors from CT scans. Furthermore, it is important to acknowledge the proposed model's limitations, which include its limited ability to segment tiny tumors.

## 5 Conclusion

The issue of accurate liver tumor detection and segmentation happens to be an open challenge in the field of medical imaging. There are a lot of gaps that exist in past literature of this domain, and it was significantly important to address those gaps. Many of such gaps deal with low-level contrast, intensity issues, irregularity of tumor shape and size, and growing stages of tumor. In continuation to the existing research, we proposed a novel deep learning model namely RA-Net for accurate and precise tumor segmentation. This model is an extended and improved version of the U-Net model. The proposed RA-Net extracts the tumor features from parallel structure-based Atrous convolutions and residual modules simultaneously. The parallel structure-based Atrous convolution grabs the broader context while extracting the features from multi-view kernels without adding any extra parameters to the network. Moreover, the choice of activation function plays an essential role in network performance. In this work, a non-monoatomic activation namely Mish is deployed. Further, the Residual model in parallel with Atrous convolutions is also employed in which tumor features are extracted by learning residual mapping. All these characteristics improve the feature learning process of the model and hence the performance increase. In addition, the proposed framework exhibits an encouraging Jaccard score of 72% on the 3DIRCADB dataset and offers a complete end-to-end solution to detect tumors directly from CT scans. In the future, we will evaluate the model on other larger tumor segmentation databases.

**Acknowledgements** This research was supported by the Chung-Ang University Research Scholarship Grants in 2021 and also supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2019R1F1A1060668).

## References

1. Organization WH (1983) Prevention of liver cancer: report of a WHO meeting [held in Geneva from 30 January to 4 February 1983]. World Health Organization
2. Davis GL, Dempster J, Meler JD, Orr DW, Walberg MW, Brown B, Berger BD, O'Connor JK, Goldstein RM (2008) Hepatocellular carcinoma: management of an increasingly common problem. In: Baylor University Medical Center Proceedings, vol 3. Taylor & Francis, pp 266–280
3. Li W (2015) Automatic segmentation of liver tumor in CT images with deep convolutional neural networks. *J Comput Commun* 3(11):146
4. Li BN, Chui CK, Chang S, Ong SH (2012) A new unified level set method for semi-automatic liver tumor segmentation on contrast-enhanced CT images. *Expert Syst Appl* 39(10):9661–9668
5. Moghbel M, Mashohor S, Mahmud R, Saripan MIB (2016) Automatic liver tumor segmentation on computed tomography for patient treatment planning and monitoring. *EXCLI J* 15:406
6. Chlebus G, Meine H, Moltz JH, Schenk A (2017) Neural network-based automatic liver tumor segmentation with random forest-based candidate filtering. arXiv preprint [arXiv:1706.00842](https://arxiv.org/abs/1706.00842)
7. Kumar S, Moni R, Rajeesh J (2011) Automatic segmentation of liver and tumor for CAD of liver. *J Adv Inf Technol* 2(1):63–70
8. Moltz JH, Bornemann L, Dicken V, Peitgen H (2008) Segmentation of liver metastases in CT scans by adaptive thresholding and morphological processing. In: MICCAI Workshop, vol 43, p 195
9. Zhao J, Li D, Kassam Z, Howey J, Chong J, Chen B, Li S (2020) Tripartite-GAN: synthesizing liver contrast-enhanced MRI to improve tumor detection. *Med Image Anal* 63:101667
10. Kumar S, Devapal D (2014) Survey on recent CAD system for liver disease diagnosis. In: 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT). IEEE, pp 763–766
11. Gruber N, Antholzer S, Jaschke W, Kremser C, Haltmeier M (2019) A joint deep learning approach for automated liver and tumor segmentation. In: 2019 13th International Conference on Sampling Theory and Applications (SampTA). IEEE, pp 1–5
12. Trivizakis E, Manikis GC, Nikiforaki K, Drevelegas K, Constantinides M, Drevelegas A, Marias K (2018) Extending 2-D convolutional neural networks to 3-D for advancing deep learning cancer classification with application to MRI liver tumor differentiation. *IEEE J Biomed Health Inform* 23(3):923–930
13. Wang P, Chen P, Yuan Y, Liu D, Huang Z, Hou X, Cottrell G (2018) Understanding convolution for semantic segmentation. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, pp 1451–1460
14. Nawaz H, Maqsood M, Afzal S, Aadil F, Mehmood I, Rho S (2020) A deep feature-based real-time system for Alzheimer disease stage detection. *Multimed Tools Appl* 80:35789–35807
15. Xia K, Yin H, Qian P, Jiang Y, Wang S (2019) Liver semantic segmentation algorithm based on improved deep adversarial networks in combination of weighted loss function on abdominal CT images. *IEEE Access* 7:96349–96358
16. Zhang J, Xie Y, Zhang P, Chen H, Xia Y, Shen C (2019) Light-weight hybrid convolutional network for liver tumor segmentation. In: *IJCAI*, pp 4271–4277
17. Habib AB, Akhter ME, Sultaan R, Zahir ZB, Arfin R, Haque F, Amir SAB, Hussain MS, Palit R (2020) Performance analysis of different 2D and 3D CNN model for liver semantic segmentation: a review. In: International Conference on Medical Imaging and Computer-Aided Diagnosis. Springer, pp 166–174
18. Kuo C-L, Cheng S-C, Lin C-L, Hsiao K-F, Lee S-H (2017) Texture-based treatment prediction by automatic liver tumor segmentation on computed tomography. In: 2017 International Conference on Computer, Information and Telecommunication Systems (CITS). IEEE, pp 128–132
19. Almotairi S, Kareem G, Aouf M, Almutairi B, Salem MA-M (2020) Liver tumor segmentation in CT scans using modified SegNet. *Sensors* 20(5):1516
20. Umer J, Irtaza A, Nida N (2020) MACCAI LiTS17 liver tumor segmentation using RetinaNet. In: 2020 IEEE 23rd International Multitopic Conference (INMIC). IEEE, pp 1–5
21. Wong D, Liu J, Fengshou Y, Tian Q, Xiong W, Zhou J, Qi Y, Han T, Venkatesh S, Wang S-C (2008) A semi-automated method for liver tumor segmentation based on 2D region growing with knowledge-based constraints. In: MICCAI Workshop, vol 43, p 159

22. Li X, Chen H, Qi X, Dou Q, Fu C-W, Heng P-A (2018) H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes. *IEEE Trans Med Imaging* 37(12):2663–2674
23. Huang W, Li N, Lin Z, Huang G-B, Zong W, Zhou J, Duan Y (2013) Liver tumor detection and segmentation using kernel-based extreme learning machine. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp 3662–3665
24. Wu W, Wu S, Zhou Z, Zhang R, Zhang Y (2017) 3D liver tumor segmentation in CT images using improved fuzzy C-means and graph cuts. *BioMed Res Int* 2017:1–11
25. Raj A, Jayasree M (2016) Automated liver tumor detection using Markov random field segmentation. *Procedia Technol* 24:1305–1310
26. Yang Z, Zhao Y, Liao M, Di S, Zeng Y (2021) Semi-automatic liver tumor segmentation with adaptive region growing and graph cuts. *Biomed Signal Process Control* 68:10267
27. Zhang X, Tian J, Deng K, Wu Y, Li X (2010) Automatic liver segmentation using a statistical shape model with optimal surface detection. *IEEE Trans Biomed Eng* 57(10):2622–2626
28. Li D, Liu L, Chen J, Li H, Yin Y, Ibragimov B, Xing L (2016) Augmenting atlas-based liver segmentation for radiotherapy treatment planning by incorporating image features proximal to the atlas contours. *Phys Med Biol* 62(1):272
29. Chartrand G, Cresson T, Chav R, Gotra A, Tang A, De Guise JA (2016) Liver segmentation on CT and MR using Laplacian mesh optimization. *IEEE Trans Biomed Eng* 64(9):2110–2121
30. Luo Q, Qin W, Wen T, Gu J, Gaio N, Chen S, Li L, Xie Y (2013) Segmentation of abdomen MR images using kernel graph cuts with shape priors. *Biomed Eng Online* 12(1):1–19
31. Christ PF, Elshaer MEA, Ettliger F, Tatavarty S, Bickel M, Bilic P, Rempfler M, Armbruster M, Hofmann F, D'Anastasi M (2016) Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp 415–423
32. Tran S-T, Cheng C-H, Liu D-G (2020) A multiple layer U-Net, U n-Net, for liver and liver tumor segmentation in CT. *IEEE Access* 9:3752–3764
33. Seo H, Huang C, Bassenne M, Xiao R, Xing L (2019) Modified U-Net (mU-Net) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in CT images. *IEEE Trans Med Imaging* 39(5):1316–1325
34. Liu Z, Song Y-Q, Sheng VS, Wang L, Jiang R, Zhang X, Yuan D (2019) Liver CT sequence segmentation based with improved U-Net and graph cut. *Expert Syst Appl* 126:54–63
35. Bai Z, Jiang H, Li S, Yao Y-D (2019) Liver tumor segmentation based on multi-scale candidate generation and fractal residual network. *IEEE Access* 7:82122–82133
36. Budak Ü, Guo Y, Tanyildizi E, Şengür A (2020) Cascaded deep convolutional encoder-decoder neural networks for efficient liver tumor segmentation. *Med Hypotheses* 134:109431
37. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp 234–241
38. Christ PF, Ettliger F, Grün F, Elshaera MEA, Lipkova J, Schlecht S, Ahmaddy F, Tatavarty S, Bickel M, Bilic P (2017) Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks. *arXiv preprint arXiv:1702.05970*
39. He K, Zhang X, Ren S, Sun J (2015) Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision, pp 1026–1034
40. Li S, Tso GKF (2018) Bottleneck supervised u-net for pixel-wise liver and tumor segmentation. *arXiv preprint arxiv:1810.10331*
41. Ali Z, Irtaza A, Maqsood M (2021) An IOMT assisted lung nodule segmentation using enhanced receptive field-based modified UNet. *Pers Ubiquitous Comput.* <https://doi.org/10.1007/s00779-021-01637-x>
42. Bukhari M, Bajwa KB, Gillani S, Maqsood M, Durrani MY, Mehmood I, Ugail H, Rho S (2020) An efficient gait recognition method for known and unknown covariate conditions. *IEEE Access* 9:6465–6477
43. Afzal S, Maqsood M, Mehmood I, Niaz MT, Seo S (2021) An efficient false-positive reduction system for cerebral microbleeds detection. *CMC Comput Mater Contin* 66(3):2301–2315
44. Alirri OI, Rahni AAA, Golkar E (2018) An automated liver tumour segmentation from abdominal CT scans for hepatic surgical planning. *Int J Comput Assist Radiol Surg* 13(8):1169–1176



45. Han Y, Li X, Wang B, Wang L (2021) Boundary loss-based 2.5 D fully convolutional neural networks approach for segmentation: a case study of the liver and tumor on computed tomography. *Algorithms* 14(5):144
46. Zhang C, Ai D, Feng C, Fan J, Song H, Yang J (2020) Dial/hybrid cascade 3DResUNet for liver and tumor segmentation. In: *Proceedings of the 2020 4th International Conference on Digital Signal Processing*, pp 92–96

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

## Authors and Affiliations

Anum Kalsoom<sup>1</sup> · Muazzam Maqsood<sup>1</sup> · Sadaf Yasmin<sup>1</sup> · Maryam Bukhari<sup>1</sup> · Zian Shin<sup>2</sup> · Seungmin Rho<sup>3</sup>

✉ Seungmin Rho  
smrho@cau.ac.kr

<sup>1</sup> Department of Computer Science, COMSATS University Islamabad, Attock Campus, Attock, Pakistan

<sup>2</sup> Department of Security Convergence Science, Chung-Ang University, Seoul, Korea

<sup>3</sup> Department of Industrial Security, Chung-Ang University, Seoul 06974, South Korea