

Article

A Residual-Learning-Based Multi-Scale Parallel-Convolutions-Assisted Efficient CAD System for Liver Tumor Detection

Muazzam Maqsood ^{1,*}, Maryam Bukhari ¹, Zeeshan Ali ², Saira Gillani ³, Irfan Mehmood ⁴, Seungmin Rho ⁵ 
and Young-Ae Jung ^{6,*} 

¹ Department of Computer Science, COMSATS University Islamabad-Attock Campus, Attock 43600, Pakistan; sp21-rcs-008@cuiatk.edu.pk

² R & D Department, National University of Computer and Emerging Sciences, Islamabad 44000, Pakistan; Zeeshan.ali@nu.edu.pk

³ Department of Computer Science, Bahria University-Lahore, Lahore 54600, Pakistan; sairagillani.bukc@bahria.edu.pk

⁴ Centre for Visual Computing, University of Bradford, Bradford BD71DP, UK; I.Mehmood4@bradford.ac.uk

⁵ Department of Industrial Security, Chung-Ang University, Seoul 06974, Korea; smrho@cau.ac.kr

⁶ Division of Information Technology Education, Sunmoon University, Asan 31460, Korea

* Correspondence: muazzam.maqsood@cuiatk.edu.pk (M.M.); yajung@sunmoon.ac.kr (Y.-A.J.)

Abstract: Smart multimedia-based medical analytics and decision-making systems are of prime importance in the healthcare sector. Liver cancer is commonly stated to be the sixth most widely diagnosed cancer and requires an early diagnosis to help with treatment planning. Liver tumors have similar intensity levels and contrast as compared to neighboring tissues. Similarly, irregular tumor shapes are another major issue that depends on the cancer stage and tumor type. Generally, liver tumor segmentation comprises two steps: the first one involves liver identification, and the second stage involves tumor segmentation. This research work performed tumor segmentation directly from a CT scan, which tends to be more difficult and important. We propose an efficient algorithm that employs multi-scale parallel convolution blocks (MPCs) and Res blocks based on residual learning. The fundamental idea of utilizing multi-scale parallel convolutions of varying filter sizes in MPCs is to extract multi-scale features for different tumor sizes. Moreover, the utilization of residual connections and residual blocks helps to extract rich features with a reduced number of parameters. Moreover, the proposed work requires no post-processing techniques to refine the segmentation. The proposed work was evaluated using the 3DIRCADb dataset and achieved a Dice score of 77.15% and 93% accuracy.

Keywords: liver tumor segmentation; smart healthcare system; residual learning; multi-scale features



Citation: Maqsood, M.; Bukhari, M.; Ali, Z.; Gillani, S.; Mehmood, I.; Rho, S.; Jung, Y.-A. A Residual-Learning-Based Multi-Scale Parallel-Convolutions-Assisted Efficient CAD System for Liver Tumor Detection. *Mathematics* **2021**, *9*, 1133. <https://doi.org/10.3390/math9101133>

Academic Editors: Darian M. Onchis, Dan Pescaru, Flavia Micota, Pedro Real and Codruta Istin

Received: 16 March 2021

Accepted: 7 May 2021

Published: 17 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Recent advancements in machine learning and computer communication help to improve smart healthcare systems, especially using IOMT devices [1,2]. A multimedia-based medical diagnostic system is one of the primary necessities in the medical healthcare sector. These intelligently designed diagnostic systems [3] provide solutions that are further helpful for radiologists and physicians [4]. Liver cancer is listed as the sixth most widespread cancer across the globe. Hepatocellular carcinoma (HCC) is the primary cancer type, causing about 700,000 deaths on a yearly basis [5]. The most observed ground cause of primary liver cancers tends to be cirrhosis. Cirrhosis is usually caused by consuming an excessive amount of alcohol, hepatitis B and C viruses, and other liver diseases, which betide because of weight gain.

Cirrhosis can be diagnosed with the help of imaging testing approaches, such as MRI, CT, or ultrasound. CT is a very well-known image testing technique, as it provides comprehensive cross-sectional abdominal images and is used for liver tumor segmentation [6].

The issue of intensity and luminosity setup continues to exist as tumor and neighboring liver tissues have similar appearances, resulting in poor contrast and ultimately affecting the performance [7]. Hence, it is mandatory to cope with this scenario, and for this, image enhancement is required for better tumor detection performance [8]. The cancer tumor can be identified using a sharp color difference in the pixel intensity by examining the CT scans. This difference is called hypodense if it appears to be dark in shade compared to the neighbor tissues inside the liver [9]. The conventional method of tumor segmentation is not well suited for a clinical examination [10]. This is because the human liver stretches over 150 slices on average in a CT scan and tumor shapes are irregular [11]. These issues make automated CAD systems highly important for liver tumor detection [12]. Presently, almost all of the traditional methods for tumor extraction are time-consuming and need expert knowledge [13].

A convolutional neural network (CNN) is one of the most widely used methods in recent years for liver tumor detection [14]. Furthermore, semantic segmentation manages the process of assigning a label to every pixel of an image and produces very good results [15]. The shape and size of the lesion is a very challenging factor in the medical imaging domain. All deep learning and semantic segmentation models acquire rich contextual data by increasing the number of convolution layers [16]. The convolution process should also focus on the contextual information, especially for medical imaging scenarios where lesions have similarities with surrounding regions, causing inaccurate segmentation [17,18]. Considering this, it is crucial to obtain detailed contextual information while investigating the ROI. Patch-based approaches have been used to manage these issues, where images are converted to small patches [19]. Moreover, liver tumor detection should begin with robust identification of the liver [20]. Only then is the basis provided for accurate tumor identification. The recent advanced CAD methods still find it difficult to detect liver tumors directly, and therefore, it remains an important research problem.

Along this line of research, semantic segmentation has been used as a baseline for liver tumor detection. We propose a novel algorithm using multi-scale parallel convolution blocks (MPCs) and Res blocks based on residual learning. The fundamental idea of utilizing multi-scale parallel convolutions of varying filter sizes in MPCs is to extract multi-scale features of different-sized tumors. Moreover, the utilization of residual connections and residual blocks is helpful in feature extraction due to the addition of feature maps provided by all previous layers to the next layer. Tumor segmentation is a much more challenging task as compared to liver segmentation. The proposed method in this research was tuned in such a way that it performs tumor segmentation directly. Despite direct tumor segmentation being a complicated process, the proposed method showed promising results as compared to already existing methods. The proposed technique was evaluated using the 3Dircadb dataset and a comparative analysis was also performed to validate the proposed model. The contributions of the proposed work are given as follows:

- A complete end-to-end algorithm for segmenting tumors directly from CT scans with no post-processing step.
- In our algorithm, we utilized the MPCs to extract multi-scale features of different-sized and -shaped tumors.
- The residual learning approach was also employed in our algorithm by using residual connections and residual blocks, which helped with extracting deep features.

The rest of this paper is organized as follows: Section 2 provides an extensive literature review, Section 3 discusses the details of our proposed method, Section 4 presents results and discussion, followed by the conclusion in Section 5.

2. Related Work

The literature presents a diverse range of techniques and methodologies for liver tumor detection and segmentation. Christ et al. [21] proposed a model for liver and lesion segmentation using cascaded deep neural networks and 3D conditional random fields. The two cascaded U-Net models were used to perform liver and tumor segmentation

and the resultant outputs were passed on to 3D conditional random fields with a DSC of 0.943 for liver segmentation. Sun et al. [22] proposed a multichannel fully convolutional network (FCN) for contrast-enhanced multiphase CT scan images. A single channel of the FCN consists of eight convolutional layers, three subsampling layers, three deconvolution layers, and two feature fusion layers. The convolutional layers used varied kernel sizes and acquired features from the image by keeping spatial correlations. All fully convolutional network channels were passed through an independent training phase and achieved a volumetric overlap error (VOE) of 8.1 ± 4.5 . Chlebus et al. [9] also proposed a modified U-Net model for liver tumor segmentation using short skip connections for parameter renewals and speed enhancement of the model. The output was subjected to post-processing using a shape-based method and achieved a DSC of 0.58 using the random forest technique.

Liu et al. [23] modified the existing research work of Christ et al. [21] and Chlebus et al. [24] with their proposed GIU-Net by combining U-Net with a graph cut algorithm. They increased the depth of structure and made skip connections from the pooling layers output, combining it with a graph-cut approach while achieving a DSC score of 0.9505. Later, Li et al. [25] contradicted Liu et al. [26] and a similar approach presented by other researchers. They primarily focused on FCN-8's structure during the segmentation phase. The proposed model in this research had four major max-pooling layers with two skip structures to merge the final two outputs of the max-pooling layer with the parallel up-sampling layer. There have been two additional skip connections for the integration of residual outputs of the max-pooling layer with a parallel up-sampling layer. The expected accuracy of 0.994 was not achieved due to the noise present in the input. Budak et al. [27] also presented two cascaded encoder–decoder CNNs for liver tumor segmentation and used an EDCNN algorithm with two symmetric encoder and decoder parts. The two parts had ten convolutional layers with batch normalization and ReLu activation, followed by a max-pooling layer. In the next step, segmentation was performed using two cascaded deep neural networks, with one focused on the liver and the other on the tumor. The output of the former network was forwarded as the input to the latter network. The DSC values of 0.9522 and 0.634 were gained in the liver and tumor segmentations, respectively.

All the previous research works employed deep-learning-based architectures for the efficient segmentation of liver tumors. It is generally observed that tumor segmentation is performed after liver ROI extraction, which requires the model to be trained in two stages. In the first stage, the model is trained to segment the liver from the whole CT scan and in the second stage, they again trained the model to extract the tumor from the extracted liver ROI. Moreover, they also pass the output of their models to post-processing techniques and methods to improve the segmentation and performance accuracy. By considering these issues, we propose a complete end-to-end segmentation model that is capable enough to segment the liver tumors directly from a CT scan and requires no need for any preprocessing techniques. In our proposed approach, there is no need to first segment the liver for tumor segmentation and our method can assist radiologists in better treatment planning by providing early and accurate tumor detection.

3. Materials and Methods

The main flow of the proposed methodology is shown in Figure 1, which includes the dataset extraction, followed by the preprocessing of the CT scan.

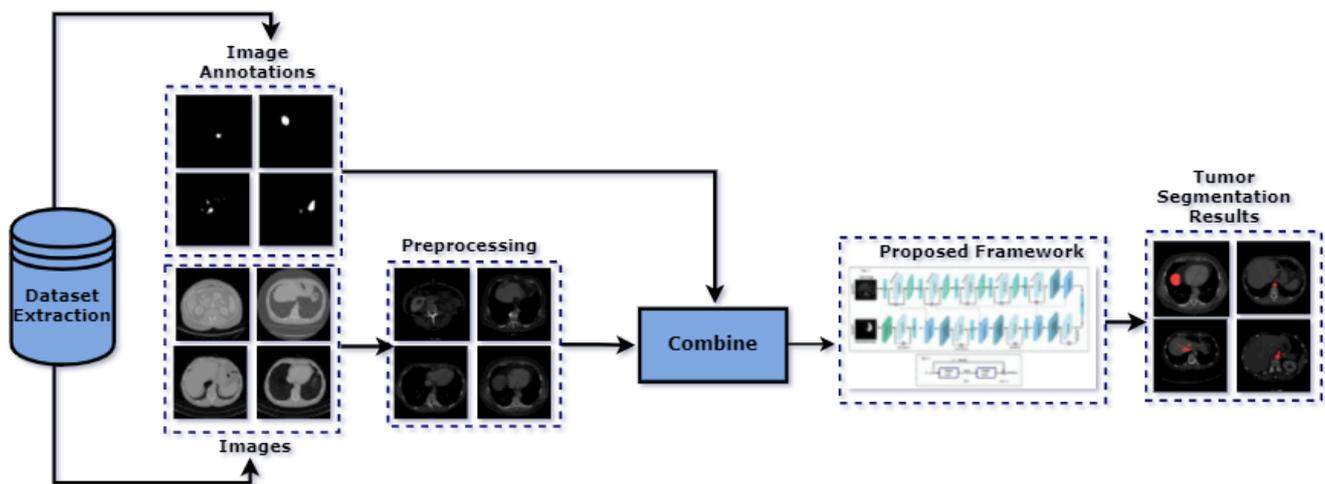


Figure 1. Schematic representation of the proposed methodology.

3.1. Dataset Extraction

This is a basic experimentation step in research work that impacts the overall system performance. In this research work, we have used the 3Dircadb dataset, which is also known as the 3D Image Reconstruction for Comparison of Algorithm Database [28]. There are a total of 20 folders with tumor CT scans from multiple European hospitals. To be more specific, the 3Dircadb dataset comprises CT scans of 20 patients diagnosed with a hepatic tumor in 75% of cases. Patient images are present in DICOM format, along with the corresponding label images and ROIs. The total number of CT slices in each 3D image varies from patient to patient. Moreover, there are some slices in which the tumor is not present, and we also consider those slices in our experimentation. The size of the 2D CT scan slice that was used to train the algorithm is $256 \times 256 \times 1$. The details of the dataset are presented in Table 1.

Table 1. Dataset details for the dataset used in this research.

S.No	Gender	YOB	Voxel Size (mm)	Image Size (pixels)	Liver Size (cm)	Average Liver Density	Segmentation Challenges
1	F	1944	0.57	512	18.3	111	Stomach, pancreas, duodenum
			0.57	512	15.1		
			1.6	129	14.1		
2	F	1987	0.78	512	20.1	84	Pancreas, duodenum
			0.78	512	16.9		
			1.6	172	15.7		
3	M	1956	0.62	512	16.7	108	Artifact due to metal
			0.62	512	14.9		
			1.25	200	15.2		
4	M	1942	0.74	512	16.9	107	Heart
			0.74	512	12.0		
			2	91	17.2		
5	M	1957	0.78	512	19.8	69	Diaphragm, duodenum
			0.78	512	16.8		
			1.6	139	19.1		

3.2. Preprocessing

Generally, medical imaging datasets have a noisy texture that causes the ROI to fade out. The noise could include any kind of blotches, irregular spots, unwanted objects, and organs. A medical imaging dataset needs to be preprocessed first to make it suitable for further experimentation. This step is mandatory to achieve enhanced images, as raw data

is noisy most of the time and cannot be processed further. It is very important to enhance ROIs by eliminating unwanted noise; for this purpose, various researchers have proposed multiple techniques. Mostly, contrast enhancement is used to improve the image quality. This is done using windowed Hounsfield unit values in the range $[-100, 400]$, which results in an enhanced image without any kind of noisy blotches, irregular spots, organs, and unwanted objects. We have used this preprocessing step over the dataset to enhance the visibility of ROI and achieve a better image quality. This preprocessing technique is also followed by other researchers [27,29]. Figure 2 shows some samples from the 3DIRCADb dataset before and after the enhancement operation was applied.

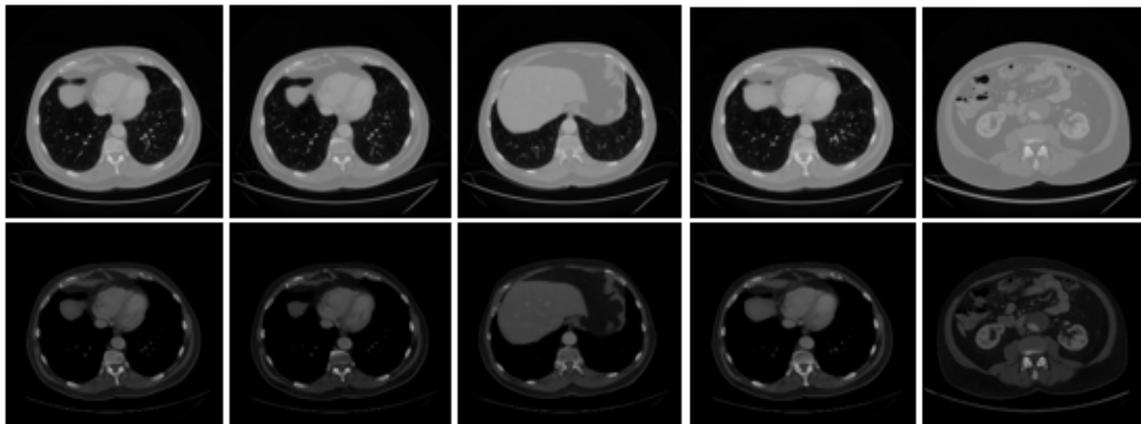


Figure 2. Samples of raw and enhanced CT scan images from the 3DIRCADb Dataset.

3.3. Architecture

In this section, we explain our proposed novel segmentation architecture for efficient liver tumor segmentation. The proposed architecture mainly consists of a down-sampling path, a bottleneck path, and an up-sampling path. Each of these paths employs the use of multi-parallel convolution blocks (MPCs) and Res blocks. The architecture of the proposed algorithm is shown in Figure 3.

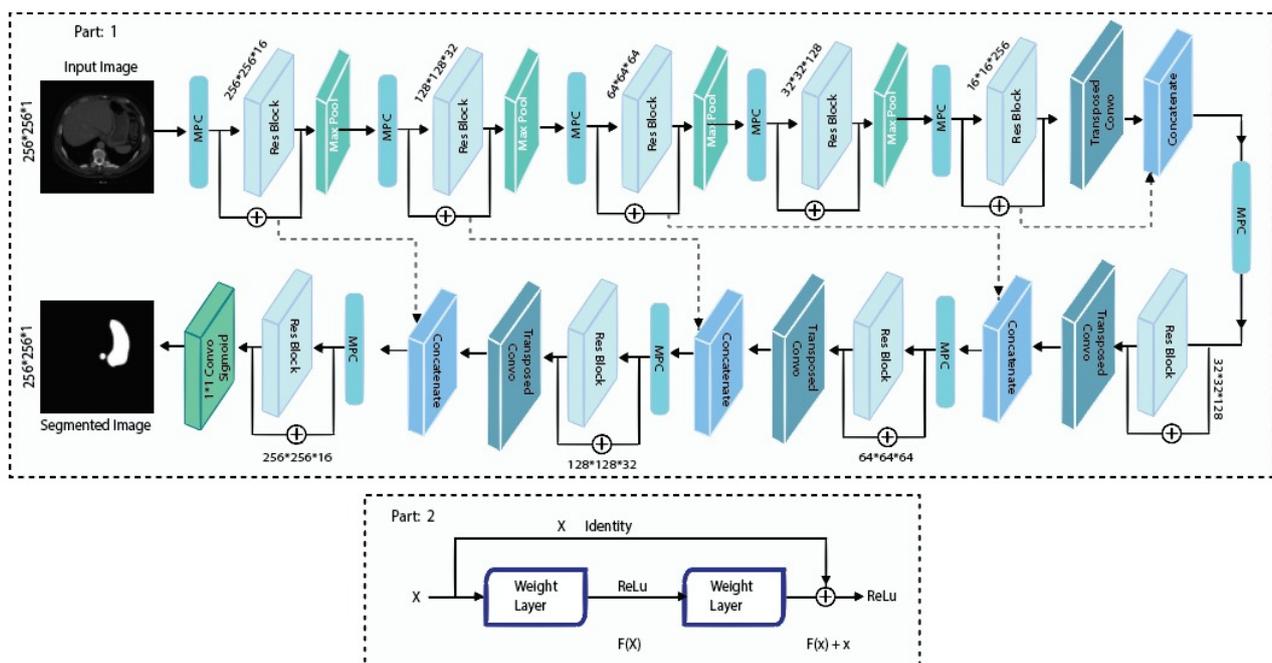


Figure 3. Architecture of the proposed algorithm for the segmentation of liver tumors.

3.3.1. Down-Sampling Layers

The down-sampling path starts by using the CT scan image with size $256 \times 256 \times 1$ as the input to the multi-scale parallel convolution block (MPC), as shown in Figure 3, followed by the Res block and max-pooling operation of size 2×2 to reduce the spatial dimensions of the given CT scan. This process is defined in Equation (1):

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

In Equation (1), a neuron $y_{k,w}^i$ is present on a position (k, w) at the i th output map of the downsampling layer. In the i th input map x_i , a neuron $y_{k,w}^i$ is assigned with a maximum value in region $p \times p$.

Multi-Scale Parallel Convolution Blocks (MPCs):

The architecture of the MPC contains parallel convolutions utilizing different filter scale sizes of 1×1 , 3×3 , and 5×5 followed by ReLu [30] activations, defined as in Equation (2):

$$f(x) = \{x \text{ if } x > 0 \text{ otherwise } 0\} \tag{2}$$

The output of these multi-parallel convolutions is added and given as an input to the Res block. The utilization of MPCs is a very powerful unit and it learns at different scales as it employs the use of parallel convolutions that are capable of extracting features of different-sized tumors. The MPC is used after every max-pool operation, except the first MPC. The feature maps of different convolutions are calculated using the following Equation (3):

$$g[x, y] = (m * n)[x, y] = \sum_f \sum_h n[f, h] f[x - f, y - h] \tag{3}$$

In the above Equation (3), the input image is denoted by m , while the kernel or filter is represented by n . The indexes of the matrix's rows and columns are denoted by x and y . Moreover, the architecture of an MPC is illustrated in Figure 4.

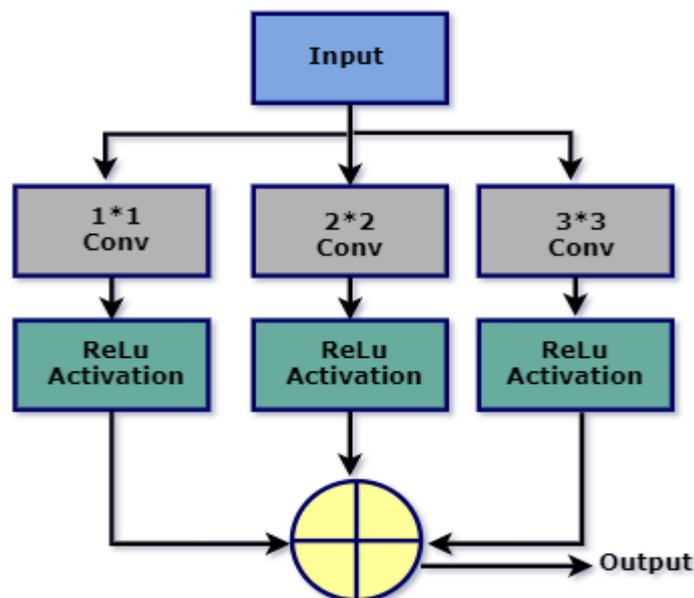


Figure 4. The architecture of an MPC.

Res Blocks:

The input of a Res block is the output of the previous MPC. The architecture of the Res block starts with a 1×1 convolution to produce N new elements on different positions of the feature map by giving input of N different elements. An element value x_i is present at the position of the i th input channel. Similarly, at the same corresponding position, the

output value x_j is used to represent the position of the j th output channel. The weight matrix between x_i and x_j is represented by w_{ij} , while the bias term is represented by b_j . Therefore, the mathematical formulation is represented by Equation (4):

$$x_i = \left(\sum_i^N w_{ij} * x_i \right) + b_j \quad (4)$$

This 1×1 convolution acts as a projection layer and decreases the number of filters or kernels at the end layer and increases them at the first layer. This approach is known as the projection shortcut utilized by [24] and can be defined as in Equation (5):

$$y = F(x, \{W_i\}) + x \quad (5)$$

The input and output layer vectors are represented by x and y . A residual mapping that is to be learned is represented by the term $F(x, \{W_i\})$. In our Res blocks, there are two layers $F = W_2\sigma(W_1x)$, as shown in Figure 3 part 2, in which the ReLU activation function is represented by the term σ . Therefore, with the help of an addition and shortcut connection, an $F + x$ operation is computed. In Equation (4), this shortcut connection does not introduce any extra parameter in the network. In Equation (5), the addition between F and x is only performed if their dimensions are equal. The shortcut connection performs a linear projection in the case when the dimensions are unequal. The linear projection is denoted by W_s and it is given by Equation (6):

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

Our down-sampling path follows the same pattern four consecutive times, followed by a max-pool and dropout layer of 0.05 to prevent overfitting of the model. The output of the multi-parallel convolution block is also added to the output of the Res block. Moreover, the filter size for each of our convolution blocks is 16, 32, 64, and 128, respectively.

3.3.2. Bottleneck Layer

The bottleneck layer of our proposed architecture consists of a multi-scale parallel convolution block (MPC) followed by Res blocks, as shown in Figure 3 part 1. The output of the last max-pooling layer in the down-sampling path is given as an input to the MPC. It contains filters of different sizes, whose outputs are added and given as the input to the Res block. The output of the bottleneck layers is given as the input to the transposed convolution layer, which is the first up-sampling layer. The size of the feature map, which is given as an input to the bottleneck layer, is $16 \times 16 \times 256$, with the total number of filters set to 32.

3.3.3. Up-Sampling Layers

The up-sampling layers utilize the transposed convolution of size 3×3 and stride of 2×2 . The transposed convolution serves as a deconvolution layer and performs the up-sampling of images with proper learning instead of the simple up operation, which only doubles the dimension of the input image without any weights. They are also known as fractionally strided convolutions. Suppose that if a convolution is applied from left to right on inputs and outputs which are unrolled into vectors by kernel w and stride of one unit without padding, then we have a matrix called a sparse matrix C through which convolution can be represented. In a sparse matrix C , the non-zero elements of the kernels are represented by W_{ij} . On the other hand, if the transpose of a sparse matrix C is obtained, then a backward pass of the convolution operation is easily attained. The error is backpropagated and the transpose of a sparse matrix is multiplied by the loss. A convolution is defined by kernel w , whose forward and backward passes are calculated by taking the product of the sparse matrix C and its transpose C^T . Similarly, the forward and backward passes of transposed convolution defined by kernel w are computed by

multiplying the sparse matrix C and $(C^T)^T$. The pattern of up-sampling layers consists of transposed convolutions followed by skip connections, MPCs, and Res blocks. The total number of filters in each transpose convolution layer is 128, 64, 32, and 16, respectively. The main purpose of these downsampling layers is to recover the size of the feature maps by adding spatial and contextual information to the segmentation image. We can transfer the contextual information from the down-sampling layers to the up-sampling layers with the help of skip connections. In the end, a convolution of size 1×1 followed by a sigmoid activation function is used to get the final segmented image of $256 \times 256 \times 1$.

3.3.4. Skip Connections

The loss of low-level information may happen during the down-sampling of the image. The skip connections are used to recover the information that is lost during down-sampling and to let the up-sampling layers retrieve the low-level features. This can be achieved via a concatenation operation between the up-sampling layers to the outputs of the down-sampling layers to combine the contextual information for localization. A dropout of the same rate is utilized after the concatenation operation, followed by multi-scale parallel convolution blocks (MPCs) and Res blocks with shortcut connections, respectively.

3.4. Training Details and Hyperparameters

The hyperparameters of our proposed model include the learning rate, which was 0.001, with weight optimizer adaptive learning optimization (Adam). It utilizes the momentum term along with stochastic gradient descent and RMSprop. The Adam updates the weights of the network using Equation (7):

$$W_t = W_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} \quad (7)$$

In the above-mentioned equation, the weights of the model are represented by W , and η represents the step size, where its value depends upon iterations of the network, while the values of \hat{m}_t and \hat{v}_t are computed using the equations mentioned below:

$$\hat{m}_t = \frac{\hat{m}_t}{1 - \beta_1^t} \quad \text{and} \quad \hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (8)$$

In the above-mentioned equations, the values of β_1 and β_2 are 0.9 and 0.999, respectively. During network training, the error between the actual values and predicted values are computed with the use of a loss function named the binary cross-entropy loss. It is defined below:

$$\text{BCE} = -\frac{1}{N} \sum_{i=1}^N y_i * \log(P(y_i)) + (1 - y_i) * \log(1 - p(y_i)) \quad (9)$$

In the above equation, BCE stands for binary cross-entropy in which y_i refers to the class of pixel predicted by the model, while $P(y_i)$ represents the probability predicted by the trained model for all pixels in the background or foreground. The proposed model was trained with 150 epochs with an input batch size of 4 and an image dimension of $256 \times 256 \times 1$.

4. Experimentation and Results

This section is divided into subheadings that provide a concise and precise description of the experimental results, their interpretation, and the experimental conclusions that can be drawn.

4.1. Performance Measures

4.1.1. Dice Similarity Coefficient

DSC, or Dice similarity coefficient, is commonly used to calculate the similarity between two samples. In this research work, this performance measure determined the overlap between two binary masks. It can be mathematically defined as the size of the overlap between two segmentations divided by the total size of the two objects. The provided range of DSC is usually from 0 (no overlap) to 1 (perfect overlap). DSC is calculated using the following equation:

$$\text{DSC} = \left(\frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \right) \quad (10)$$

4.1.2. Jaccard Similarity Coefficient

JSC gives segmented image and binary mask values precisely. It is also defined as the ratio of similarity and diversity of samples used in experimentation. In mathematical terms, it is the ratio of the intersection between two binary masks with their union. JSC is calculated according to the equation given below:

$$\text{JSC} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (11)$$

4.1.3. Accuracy

Accuracy is one of the most significant performance measures that determine the efficiency and effectiveness of any model. Accuracy represents the ratio of correctly segmented samples to a total number of samples [31].

4.1.4. Symmetric Volume Difference

SVD provides the difference of the segmented images from the ground truth. If the value of SVD is zero, it represents a promising resultant segmentation value. The equation determines how to calculate SVD, where DSC is the Dice similarity coefficient:

$$\text{SVD} = (1 - \text{DSC}) \quad (12)$$

4.1.5. Sensitivity

The correctly identified proportion of true positives is measured using sensitivity [32].

4.1.6. Specificity

The correctly identified proportion of true negatives is measured using specificity [32].

4.1.7. Matthew's Correlation Coefficient (MCC)

MCC is widely used for classification problems when the classes are highly imbalanced. It is also known as the "phi coefficient" and it is defined using Equation (13):

$$\text{MCC} = \frac{\text{TP} * \text{TN} - \text{FP} * \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}} \quad (13)$$

In all the above equations, true positive (TP) represents pixels that form the foreground and are classified as foreground. True negative (TN) represents pixels that form the background and are classified as background. False negative (FN) represents foreground pixels that have been inaccurately classified by the classifier as background pixels. False positive (FP) represents background pixels that have been incorrectly classified as foreground pixels by the classifier.

4.2. Results and Discussions

In this proposed work, a novel deep learning algorithm was used to segment the liver tumors directly from the CT scan. The proposed model was validated on preprocessed images of the 3DIRCADb dataset. The whole dataset of 20 patients was divided into non-overlapping train and test sets with a random 80–20 split division. Moreover, the same experiment was executed ten times and the average results are reported to avoid any bias. Table 2 shows the results of our proposed model for segmenting the liver tumors. The Dice score achieved for our proposed model was 77.15% and the Jaccard score was 68.5%. The standard deviations are also given in Table 2.

Table 2. Means \pm standard deviations of the segmentation results of proposed framework.

Authors	Dice Score	Jaccard	Accuracy	Specificity	Sensitivity	MCC	SVD
U-Net [33]	67.5 \pm 30.8%	56.0 \pm 30.7%	92 \pm 3.8%	70.1 \pm 29.6%	64.8 \pm 32.2%	0.69 \pm 29.9	0.33
Proposed Method	77.11 \pm 27.0%	67.8 \pm 26.9%	93 \pm 3.7%	79.16 \pm 20.56%	76.03 \pm 24.56%	0.766 \pm 26.06	0.23

The other evaluation metrics, which include accuracy and SVD, were also calculated. The SVD shows the difference between the actual and predicted masks. The proposed model achieved an accuracy of 93% with an SVD score of 0.23, which was the very minimum difference between the actual and predicted masks, as shown in Table 2. The reason for the higher accuracy was a class imbalance. In a given CT scan image, more pixels belong to the background class, while the number of pixels where the tumor is present much fewer. Therefore, the accuracy value is biased toward the background class because accuracy counts the total number of TP, FP, TN, and FN of all classes. The Dice and Jaccard analysis scores more accurately represent the segmentation model capability. Moreover, the values of sensitivity, specificity, and the MCC were 76.5%, 79.56%, and 0.77 respectively. We also drew a comparison of our model with U-Net proposed by Ronneberger et al. [33]. The standard U-Net architecture is very famous for biomedical image segmentation and is extensively used by different researchers. The Dice and Jaccard scores achieved by U-Net on the 3DIRCADB dataset were 67.5% and 56%, which were very low in comparison with our algorithm. The other scores achieved by U-Net included sensitivity, specificity, and the MCC, which were 70.1%, 64.8%, and 0.69 respectively. The difference between the actual and predicted masks segmented by standard U-Net was 0.33, which is an SVD score. Our model found a 9.65% improvement in the Dice score and a 12.5% improvement in the Jaccard score.

Moreover, during training, the input image of the CT scan is passed through different layers of the model, which includes convolution layers and pooling layers. The output of each layer takes the form of activation maps. The visualization of those activation feature maps of some intermediate layers of our proposed model is also shown in Figure 5. These visualizations show how the model depicts the contextual information of the image layer by layer.

Moreover, some sample images that were segmented by our model are shown in Figure 6. In Figure 6, column (A) shows the original test set slice images, column (B) shows the actual masks of the tumor, column (C) shows the actual overlay results of the original mask over the image, column (D) shows the masks that were predicted by the model, while column (E) shows the predicted overlay results of a predicted mask on the images. It is observed from Figure 6 that the model found difficulty in segmenting very small tumors, as shown in the last row images. Similarly, the model also found difficulty in segmenting in the second row, as shown in columns (B) and (D) of Figure 6. All the scores were calculated with the help of predicted and actual masks, which are shown in columns (B) and (D) of Figure 6. Furthermore, the Dice and Jaccard scores of individual CT scan slices in the test set is shown in Figure 7. In Figure 7, the x -axis shows the number of CT scan slices, while

the y -axis shows the dice and Jaccard score. It is observed from Figure 7 that, for most CT scans, the Dice score was above 80.

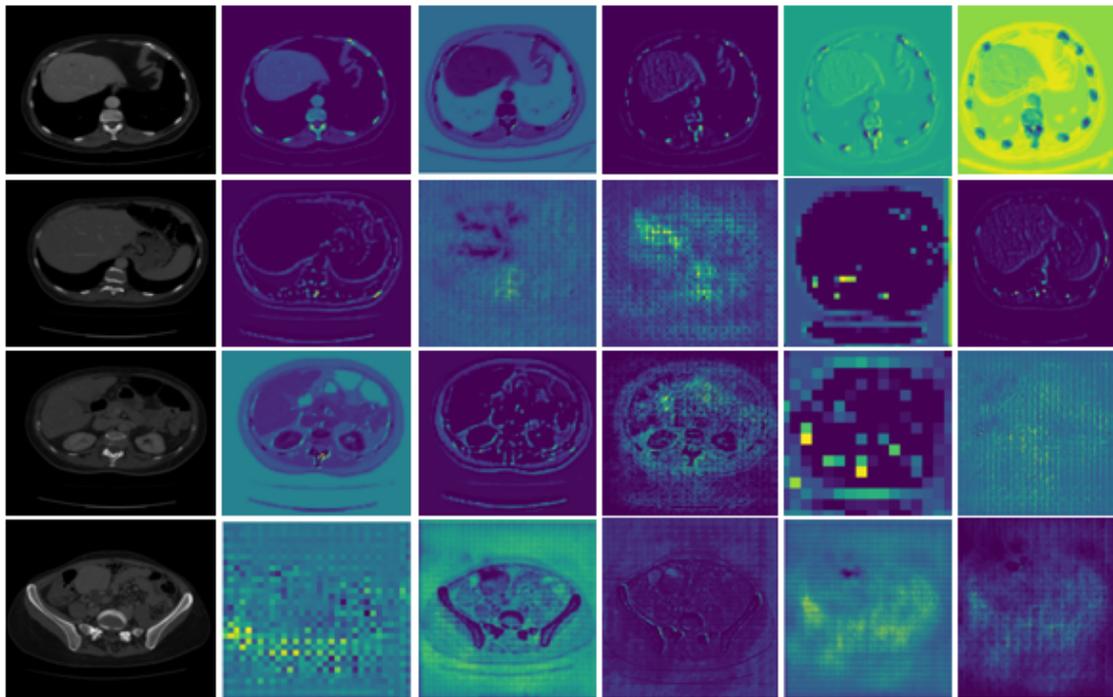


Figure 5. Visualization of the activation maps of an intermediate layer.

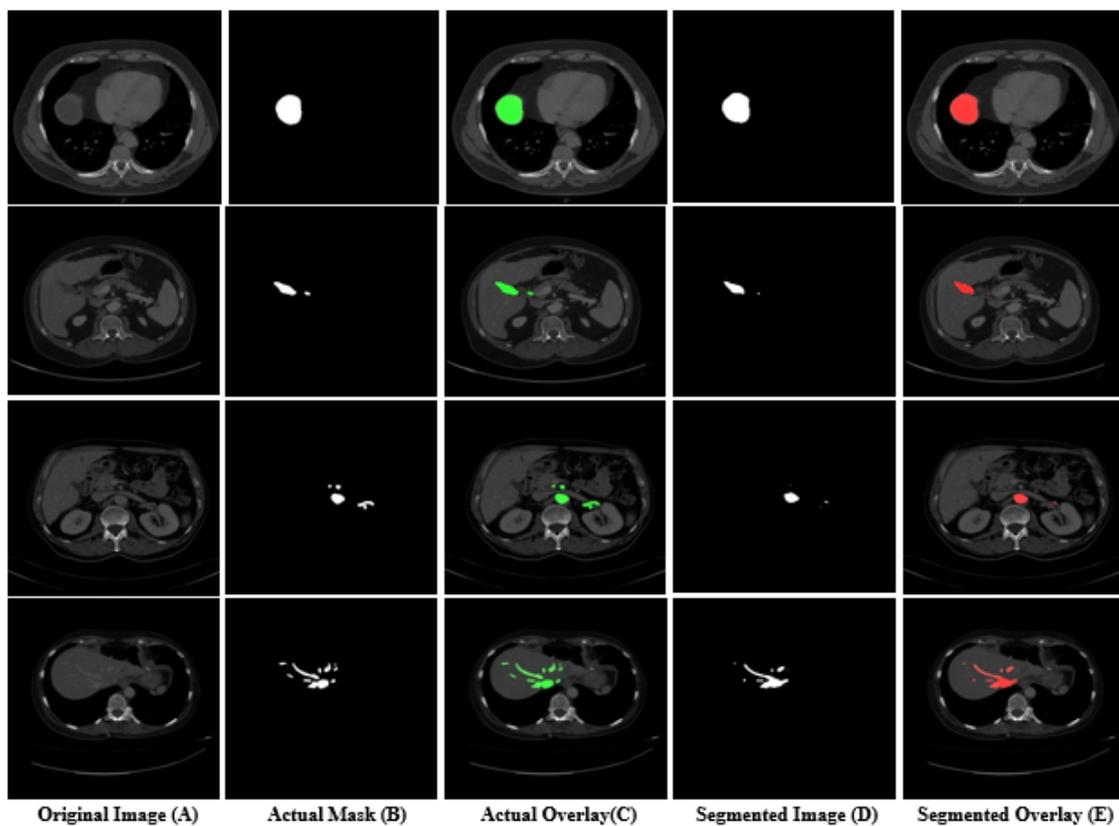


Figure 6. Segmentation results of liver tumor from a CT scan using our proposed framework.

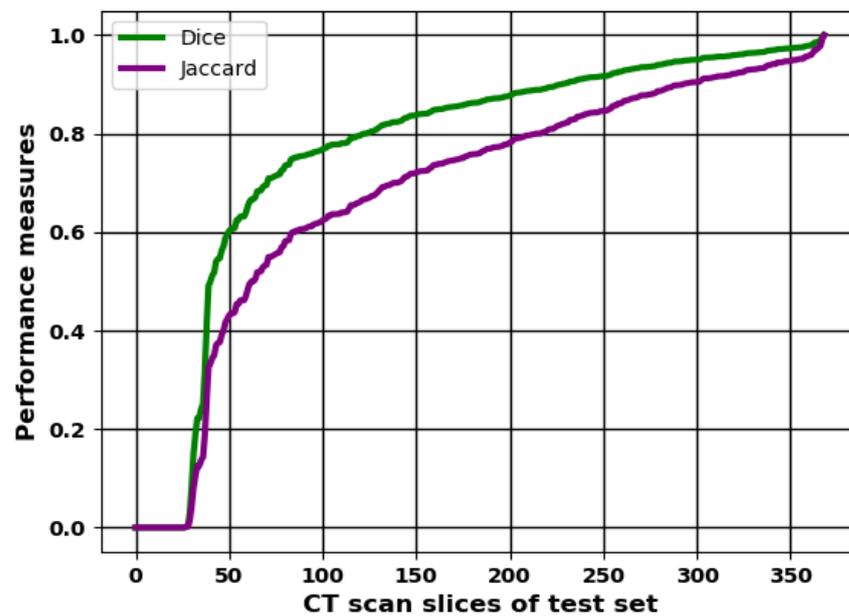


Figure 7. Analysis of the Dice and Jaccard scores over each CT scan slice of patients in the test set.

Moreover, we also checked the model loss over several epochs during the training of our proposed algorithm. Usually, when the model loss is near to zero or becomes constant over a certain number of epochs, then the model prediction is perfect. The loss curves of our proposed method are shown in Figure 8. Furthermore, the accuracy of the proposed model over epochs during training is also shown in Figure 8. Accuracy determines the correct number of predictions over all classes. The loss and accuracy curves of U-Net proposed by Ronneberger et al. [33] are also given in Figure 8.

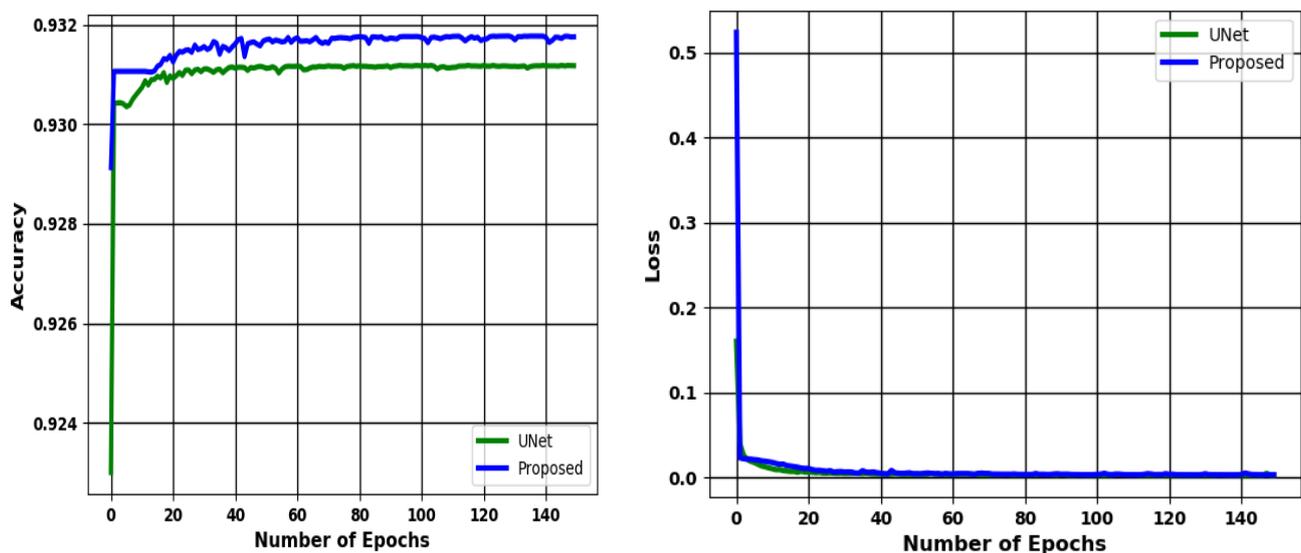


Figure 8. Accuracy and loss curves of the proposed framework and U-NET.

4.3. Comparison with State-of-the-Art Approaches

The performance and results of our proposed approach are explained in the previous section. The use of an MPC gives multi-scale features, which were shown to be very beneficial for encoding information of different-sized tumors. To compare the performance of our proposed model with existing methods, a detailed comparative analysis was performed with existing methods. It was found from the literature that by using the

3DIRCADb dataset, Christ et al. [29] achieved a 61% dice score using their proposed two FCNs in a cascaded manner. Alirr et al. [34] achieved a Dice score of 75% by utilizing the traditional method of adaptive thresholding to extract masks of liver tumors. Li et al. [35] and Z. Bai et al. [36] achieved Dice scores of 65% and 76.5%, respectively, by making some improvements upon the standard U-Net. Z. Bai et al. [36] also used an active contour model (ACM) to refine the tumors segmentation. Similarly, Budak et al. [37] achieved a Dice score of 63.4%

Moreover, by looking into the recent work for liver tumors, S.-T. Tran et al. [38] proposed an improved U-NET based method by employing the architecture of dense and dilated convolution and achieved a very significant improvement regarding the Dice score. Similarly, H. Seo et al. [39] also proposed an improved U-NET that was based on segmenting both the liver and tumors. It was observed from the previous literature that the proposed model achieved a significant improvement in segmenting tumors directly from CT scans. The main reason behind the improvement was the utilization of MPCs and the concept of residual learning to obtain features without increasing the number of parameters in the network. The multi-scale features extracted using the MPC are added with the features maps of the Res blocks to better describe the tumor features. Furthermore, the previous work in the literature developed two-stage algorithms via first segmenting the liver, followed by a liver tumor segmentation. The previous methods use post-processing techniques to refine the tumor segmentation. It is necessary to mention here that the proposed approach follows an end-to-end mechanism to segment tumors in an efficient manner. The comparison results of previous techniques, standard U-Net, and our proposed approach are given in Table 3 in terms of Dice and SVD scores.

Table 3. Comparison with the state-of-the-art approaches.

Authors	Dice Score	SVD	Year
Budak et al. [27]	63.4%	0.37	2019
Christ et al. [29]	61%	0.39	2017
Alirr et al. [32]	74.96%	0.25	2018
S.-T. Tran et al. [36]	73.34%	0.26	2020
H.seo et al. [37]	68.14%	0.32	2020
Li et al. [35]	65%	0.35	2018
Z. Bai et al. [34]	76.4%	0.24	2019
U-Net [33]	67.5%	0.33	2015
Proposed Method	77.11%	0.23	2021

5. Conclusions

This research work highlighted problems in liver tumor segmentation and provided a solution to address those issues. Many researchers have previously presented two-step methods that carry out liver segmentation, followed by tumor segmentation. This is a time-consuming approach with a higher chance of inaccuracy. To solve these issues, we proposed a technique in this research work that is challenging as it directly segments out the tumor from a CT scan. However, it is an end-to-end segmentation algorithm and efficiently performs on the sample data and provides accurate results. The proposed work employs MPCs to encode multi-scale features of different tumor sizes. The incorporation of Res blocks is also helpful for encoding tumor features with a reduced set of parameters in the network. All of these characteristics increase the segmentation performance of our model. Moreover, our approach does not require post-preprocessing steps for the refinement of segmentation results. The proposed system was evaluated using the publicly available 3DIRCADb dataset and achieved excellent results as compared to existing published work. To ensure the validity of the proposed framework, we performed a comparative analysis with already existing techniques for liver tumor detection. In the future, we will apply attention gates to our model to further improve the performance.

Author Contributions: Conceptualization, I.M.; formal analysis, M.M., M.B., and Z.A.; investigation, M.B., Z.A., and S.G.; methodology, M.M., S.G., and S.R.; validation, Y.-A.J.; visualization, Y.-A.J.; writing—original draft, I.M.; writing—review and editing, S.R. and Y.-A.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2021-2018-0-01799) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation) and also supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2019R1F1A1060668).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Bhavsar, K.A.; Abugabah, A.; Singla, J.; AlZubi, A.-A.; Bashir, A.-K. A Comprehensive Review on Medical Diagnosis Using Machine Learning. *Comput. Mater. Contin.* **2021**, *67*, 1997–2014. [[CrossRef](#)]
- Kutia, S.; Chauhdary, S.H.; Iwendi, C.; Liu, L.; Yong, W.; Bashir, A.K. Socio-Technological factors affecting user’s adoption of eHealth functionalities: A case study of China and Ukraine eHealth systems. *IEEE Access* **2019**, *7*, 90777–90788. [[CrossRef](#)]
- Tsafack, N.; Sankar, S.; Abd-El-Atty, B.; Kengne, J.; Jithin, K.; Belazi, A.; Mehmood, I.; Bashir, A.K.; Song, O.-Y.; Abd El-Latif, A.A. A new chaotic map with dynamic analysis and encryption application in internet of health things. *IEEE Access* **2020**, *8*, 137731–137744. [[CrossRef](#)]
- Bhavsar, K.A.; Singla, J.; Al-Otaibi, Y.D.; Song, O.-Y.; Zikriya, Y.B.; Bashir, A.K. Medical Diagnosis Using Machine Learning: A Statistical Review. *CMC Comput. Mater. Contin.* **2021**, *67*, 107–125. [[CrossRef](#)]
- Davis, G.L.; Dempster, J.; Meler, J.D.; Orr, D.W.; Walberg, M.W.; Brown, B.; Berger, B.D.; O’Connor, J.K.; Goldstein, R.M. Hepatocellular carcinoma: Management of an increasingly common problem. *Proc. Bayl. Univ. Med. Cent.* **2008**, *21*, 266–280. [[CrossRef](#)] [[PubMed](#)]
- Li, W.; Jia, F.; Hu, Q. Automatic Segmentation of Liver Tumor in CT Images with Deep Convolutional Neural Networks. *J. Comput. Commun.* **2015**, *3*, 6. [[CrossRef](#)]
- Li, B.N.; Chui, C.K.; Chang, S.; Ong, S.H. A new unified level set method for semi-automatic liver tumor segmentation on contrast-enhanced CT images. *Expert Syst. Appl.* **2012**, *39*, 9661–9668. [[CrossRef](#)]
- Moghbel, M.; Mashohor, S.; Mahmud, R.; Saripan, M.I.B. Automatic liver tumor segmentation on computed tomography for patient treatment planning and monitoring. *EXCLI J.* **2016**, *15*, 406.
- Chlebus, G.; Meine, H.; Moltz, J.H.; Schenk, A. Neural network-based automatic liver tumor segmentation with random forest-based candidate filtering. *arXiv* **2017**, arXiv:1706.00842.
- Kumar, S.; Moni, R.; Rajeesh, J. Automatic segmentation of liver and tumor for CAD of liver. *J. Adv. Inf. Technol.* **2011**, *2*, 63–70. [[CrossRef](#)]
- Moltz, J.H.; Bornemann, L.; Dicken, V.; Peitgen, H. Segmentation of liver metastases in CT scans by adaptive thresholding and morphological processing. In Proceedings of the MICCAI workshop, Bremen, Germany, 8 November 2008; p. 195.
- Kumar, S.; Devapal, D. Survey on recent CAD system for liver disease diagnosis. In Proceedings of the 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), Kanyakumari, India, 10–11 July 2014; pp. 763–766.
- Gruber, N.; Antholzer, S.; Jaschke, W.; Kremser, C.; Haltmeier, M. A joint deep learning approach for automated liver and tumor segmentation. In Proceedings of the 2019 13th International conference on Sampling Theory and Applications (SampTA), Bordeaux, France, 8–12 July 2019; pp. 1–5.
- Trivizakis, E.; Manikis, G.C.; Nikiforaki, K.; Drevelegas, K.; Constantinides, M.; Drevelegas, A.; Marias, K. 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.* **2018**, *23*, 923–930. [[CrossRef](#)] [[PubMed](#)]
- Wang, Y.; Wang, L.; Wang, H.; Li, P. Information-Compensated Downsampling for Image Super-Resolution. *IEEE Signal. Process. Lett.* **2018**, *25*, 685–689. [[CrossRef](#)]
- Xia, K.; Yin, H.; Qian, P.; Jiang, Y.; Wang, S. Liver semantic segmentation algorithm based on improved deep adversarial networks in combination of weighted loss function on abdominal CT images. *IEEE Access* **2019**, *7*, 96349–96358. [[CrossRef](#)]
- Zhang, J.; Xie, Y.; Zhang, P.; Chen, H.; Xia, Y.; Shen, C. Light-Weight Hybrid Convolutional Network for Liver Tumor Segmentation. In Proceedings of the IJCAI, Macao, China, 10–16 August 2019; pp. 4271–4277.
- Habib, A.B.; Akhter, M.E.; Sultaan, R.; Zahir, Z.B.; Arfin, R.; Haque, F.; Amir, S.A.B.; Hussain, M.S.; Palit, R. Performance Analysis of Different 2D and 3D CNN Model for Liver Semantic Segmentation: A Review. In Proceedings of the International Conference on Medical Imaging and Computer-Aided Diagnosis, Oxford UK, 20–21 January 2020; pp. 166–174.
- Kuo, C.-L.; Cheng, S.-C.; Lin, C.-L.; Hsiao, K.-F.; Lee, S.-H. Texture-based treatment prediction by automatic liver tumor segmentation on computed tomography. In Proceedings of the 2017 International Conference on Computer, Information and Telecommunication Systems (CITS), Dalian, China, 21–23 July 2017; pp. 128–132.

20. Wong, D.; Liu, J.; Fengshou, Y.; Tian, Q.; Xiong, W.; Zhou, J.; Qi, Y.; Han, T.; Venkatesh, S.; Wang, S.-C. A semi-automated method for liver tumor segmentation based on 2D region growing with knowledge-based constraints. In Proceedings of the MICCAI workshop, New York, NY, USA, 6–10 September 2008; p. 159.
21. Christ, P.F.; Elshaer, M.E.A.; Ettliger, F.; Tatavarty, S.; Bickel, M.; Bilic, P.; Rempfler, M.; Armbruster, M.; Hofmann, F.; D’Anastasi, M.; et al. *Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields*; Springer: Cham, Switzerland, 2016; pp. 415–423.
22. Sun, Z.; Jin, L.; Xie, Z.; Feng, Z.; Zhang, S. Convolutional multi-directional recurrent network for offline handwritten text recognition. In Proceedings of the 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), Shenzhen, China, 23–26 October 2016; pp. 240–245.
23. Liu, Z.; Song, Y.-Q.; Sheng, V.S.; Wang, L.; Jiang, R.; Zhang, X.; Yuan, D. Liver CT sequence segmentation based with improved U-Net and graph cut. *Expert Syst. Appl.* **2019**, *126*, 54–63. [[CrossRef](#)]
24. Chlebus, G.; Schenk, A.; Moltz, J.H.; van Ginneken, B.; Hahn, H.K.; Meine, H. Automatic liver tumor segmentation in CT with fully convolutional neural networks and object-based postprocessing. *Sci. Rep.* **2018**, *8*, 15497. [[CrossRef](#)] [[PubMed](#)]
25. Zhang, Z.; Wu, C.; Coleman, S.; Kerr, D. DENSE-INception U-net for medical image segmentation. *Comput. Methods Programs Biomed.* **2020**, *192*, 105395. [[CrossRef](#)] [[PubMed](#)]
26. Liu, L.; Wu, F.; Wang, Y.-P.; Wang, J. Multi-Receptive-Field CNN for Semantic Segmentation of Medical Images. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 3215–3225. [[CrossRef](#)] [[PubMed](#)]
27. Budak, Ü.; Guo, Y.; Tanyildizi, E.; Şengür, A. Cascaded deep convolutional encoder-decoder neural networks for efficient liver tumor segmentation. *Med. Hypotheses* **2020**, *134*, 109431. [[CrossRef](#)]
28. Soler, L.; Hostettler, A.; Agnus, V.; Charnoz, A.; Fasquel, J.; Moreau, J.; Osswald, A.; Bouhadjar, M.; Marescaux, J. 3D Image Reconstruction for Comparison of Algorithm Database: A Patient Specific Anatomical and Medical Image database. *IRCAD Strasbg. Fr. Tech. Rep.* **2010**. Available online: <http://www-sop.inria.fr/geometrica/events/wam/abstract-ircad.pdf> (accessed on 14 May 2021).
29. Christ, P.F.; Ettliger, F.; Grün, F.; Elshaera, M.E.A.; Lipkova, J.; Schlecht, S.; Ahmaddy, F.; Tatavarty, S.; Bickel, M.; Bilic, P. Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks. *arXiv* **2017**, arXiv:1702.05970.
30. Nair, V.; Hinton, G.E. Rectified linear units improve restricted boltzmann machines. In Proceedings of the Icm1, Haifa, Israel, 21–24 June 2010; Available online: [https://www.cs.toronto.edu/~\[h\]inton/absps/reluICML.pdf](https://www.cs.toronto.edu/~[h]inton/absps/reluICML.pdf) (accessed on 15 May 2021).
31. Afzal, S.; Maqsood, M.; Nazir, F.; Khan, U.; Aadil, F.; Awan, K.M.; Mehmood, I.; Song, O.-Y. A data augmentation-based framework to handle class imbalance problem for Alzheimer’s stage detection. *IEEE Access* **2019**, *7*, 115528–115539. [[CrossRef](#)]
32. Afzal, S.; Maqsood, M.; Mehmood, I.; Niaz, M.T.; Seo, S. An Efficient False-Positive Reduction System for Cerebral Microbleeds Detection. *CMC-Comput. Mater. Contin.* **2021**, *66*, 2301–2315. [[CrossRef](#)]
33. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
34. Alir, O.I.; Rahni, A.A.A.; Golkar, E. An automated liver tumour segmentation from abdominal CT scans for hepatic surgical planning. *Int. J. Comput. Assist. Radiol. Surg.* **2018**, *13*, 1169–1176. [[CrossRef](#)] [[PubMed](#)]
35. Li, X.; Chen, H.; Qi, X.; Dou, Q.; Fu, C.-W.; Heng, P.-A. H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes. *IEEE Trans. Med. Imaging* **2018**, *37*, 2663–2674. [[CrossRef](#)] [[PubMed](#)]
36. Bai, Z.; Jiang, H.; Li, S.; Yao, Y.-D. Liver tumor segmentation based on multi-scale candidate generation and fractal residual network. *IEEE Access* **2019**, *7*, 82122–82133. [[CrossRef](#)]
37. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
38. Tran, S.-T.; Cheng, C.-H.; Liu, D.-G. A Multiple Layer U-Net, Un-Net, for Liver and Liver Tumor Segmentation in CT. *IEEE Access* **2020**, *9*, 3752–3764. [[CrossRef](#)]
39. Seo, H.; Huang, C.; Bassenne, M.; Xiao, R.; Xing, L. 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* **2020**, *39*, 1316–1325. [[CrossRef](#)]