Journal of Digital Imaging (2023) 36:2164–2178
https://doi.org/10.1007/s10278-023-00874-1

Residual Deformable Split Channel and Spatial U-Net for Automated Liver and Liver Tumour Segmentation

S Saumiya[1] **· S Wilfred Franklin**[2]

Received: 10 February 2023 / Revised: 20 June 2023 / Accepted: 21 June 2023 / Published online: 18 July 2023

Abstract

Accurate segmentation of the liver and liver tumour (LT) is challenging due to its hazy boundaries and large shape variability. Although using U-Net for liver and LT segmentation achieves better results than manual segmentation, it loses spatial and channel features during segmentation, leading to inaccurate liver and LT segmentation. A residual deformable split depth-wise separable U-Net (RDSDSU-Net) is proposed to increase the accuracy of liver and LT segmentation. The residual deformable convolution layer (DCL) with deformable pooling (DP) is used in the encoder as an attention mechanism to adaptively extract liver and LT shape and position characteristics. Afterward, a convolutional spatial and channel features split graph network (CSCFSG-Net) is introduced in the middle processing layer to improve the expression capability of the liver and LT features by capturing spatial and channel features separately and to extract global contextual liver and LT information from spatial and channel features. Sub-pixel convolutions (SPC) are used in the decoder section to prevent the segmentation results from having a chequerboard artefact effect. Also, the residual deformable encoder features are combined with the decoder through summation to avoid increasing the number of feature maps (FM). Finally, the efficiency of the RDSDSU-Net is evaluated on the 3DIRCADb and LiTS datasets. The DICE score of the proposed RDSDSU-Net achieved 98.21% for liver segmentation and 93.25% for LT segmentation on 3DIRCADb. The experimental outcomes illustrate that the proposed RDSDSU-Net model achieved better segmentation results than the existing techniques.

Keywords U-Net · Sub-pixel convolution · Liver tumour segmentation · Deep learning · Liver segmentation

Introduction

One of the common and deadly diseases in the world is liver cancer, which poses a severe danger to human health and life [1]. The liver is commonly the site of primary or secondary tumour growth accurately segmenting the liver and LT allows doctors to analyse the function of the liver and make a diagnosis and treatment plan for the illness. However, the liver has a similar density with their surrounding organs, and LTs exhibit severe intensity inhomogeneities and low contrast in CT liver images. So it is challenging to differentiate between the margins of the liver and LT using just human eyesight [2]. Manual identification of liver and LT locations is time-consuming and less accurate. As a result, there has been an increase in the development of semi-automatic or completely automatic methods for liver and LT segmentation [3]. Prior to the development of deep learning (DL) techniques, image segmentation algorithms, like statistical models, shape region growing, and graph cut, were frequently used for semi-automatic segmentation of the LT and liver [4]. These strategies are divided into three different types: pixel-, graph-, and contour-based. Region merging and thresholding are the main components of the first technique. Pixel-based segmentation involves grouping pixels based on their characteristics. Graph-based segmentation views the image as a graph and partitions it into regions. Contour-based segmentation focuses on extracting object boundaries. However, this kind of technique can only segment the liver and LT with low accuracy [5].

Since graph-based techniques outperform pixel-based approaches in terms of performance, by simply identifying the background and foreground, this type of semi-automatic technique can accomplish precise liver segmentation without

✉ S Saumiya
saumi2424@gmail.com

S Wilfred Franklin
swfranklin@gmail.com

[1] Department of ECE, Bethlahem Institute of Engineering, Karungal, Tamil Nadu, India

[2] Department of ECE, CSI Institute of Technology, Thovalai, Tamil Nadu, India

even requiring the iterative procedure [6]. However, high computational cost occurs in this model [7]. Additionally, segmentation outcomes are affected easily by labelling results. Therefore, to increase efficiency for segmentation, researchers frequently combine graph cuts with other methods, such as watershed [8]. Contour-based LT segmentation is of more interest than the previous two types of techniques because it can produce superior segmentation results by utilising shape evolution [9].

Another approach for shape-based segmentation is the statistical shape model. This approach, unlike the level set, frequently builds a training set of liver, and LT shapes first before using machine learning techniques to create an efficient classifier [10]. These approaches have the benefit of being able to segment data more accurately than unsupervised methods [11], but they also have the drawback of requiring careful consideration when choosing a training set and classifiers [12]. Convolutional neural network (CNN)-based DL algorithms have gained attention in recent years for their use in visual recognition due to their robust nonlinear feature extraction capabilities using a variety of filters and their capacity to process large amounts of data [13]. The researcher utilised the same backbone architecture used by CNN to successfully complete an efficient classification task when addressing a semantic segmentation problem. Researchers have recently preferred to employ fully CNN (FCN) for image semantic segmentation. The decoder and encoder in these networks often use multi-level encoder-decoder topologies, and they frequently consist of a sizable number of conventional convolutional or de-convolutional layers [14].

U-Net has achieved excellent success in medical image segmentation compared to FCN. Since then, many researches have concentrated on enhancing U-Net. The most popular method is to substitute the encoder by using the backbone of traditional CNN with pre-trained networks in order to achieve transfer learning [15]. Also, U-Net is modified by adding attention mechanisms (such as attention U-Net) between decoders and encoders to concentrate on important areas. The U-Net now has additional recurrent convolution, allowing it to repeatedly extract features from the same layer. In order to prevent the rough merging of high and low features, U-Nets with various depths have also been created in place of long-range connections. Compared to 3D networks, these networks have better performance and use less memory [16].

Although liver segmentation accuracy has increased year by year due to the introduction of DL in LT, there are still some challenges in those DL models. Many of the existing approaches use convolutional layers of fixed kernel size, making it difficult for them to correctly capture liver and LT that have varying forms in various slices [17]. Additionally, pooling in existing methods causes spatial information loss, which reduces the accuracy of LT segmentation [18]. Multi-scale fusion is used in some methods to avoid this spatial information loss [19]. However, these techniques take a lot of memory and have several parameters [20]. Furthermore, some techniques simply take into account local aspects and ignore global features, which are essential for extracting neighbouring organ features to precisely segment the liver and LTs [21].

To address these challenges of accurate segmentation of the liver and LTs, a novel architecture called the residual deformable split depth-wise separable U-Net (RDSDSU-Net) is proposed. This method leverages a residual deformable encoder to accurately capture the complex shapes of the liver and LTs in medical images. In medical image segmentation, the goal is to accurately identify and delineate different anatomical structures or tumours from medical images. Medical images contain complex spatial and channel information that can be used to differentiate between different object classes or regions of interest. Splitting channel and spatial information can be beneficial in the liver and liver tumour segmentation tasks for the following reasons: First, the liver and tumour tissues may have different appearances and textures, which can be better captured by separate pathways for spatial and channel features. Second, the liver and tumour may have different shapes and sizes, and by extracting spatial features separately, the model can better capture the spatial relationships between different regions of interest. Motivated by this, CSCFSGnet is introduced to improve the liver and LT expression abilities by initially splitting channel and spatial features and then capturing the global context information from those features. Finally, in the encoder part, SPC is used to avoid the chequerboard artefact effect. Also, instead of a direct concatenation of residual deformable encoder features, the decoder summation combines the features to reduce the parameters by not increasing FMs.

The contributions of the RDSDSU-Net are described as follows:

- DCL with residual design and DP is used to build the RDSDSU-Net's encoder. It extracts the various shapes and positions of liver and LTs accurately. It enhances the generalisation ability of the proposed model.
- CSCFSGnet is introduced as the middle processing unit to enhance the feature representation. Both the spatial and channel features are extracted separately using depth-wise convolution (DWC) and point-wise convolution (PWC), respectively, to improve the liver and LT features. GCN captures global contextual liver and LT information in the channel and spatial features.
- SPC is used instead of an upsampling layer in the decoder to improve the segmentation accuracy by avoiding the chequerboard artefacts effect.

- The outcomes of the RDSDSU-Net model on the LiTS and 3DIRCADb datasets are compared with existing techniques. It obtains more accurate segmentation than the existing approaches.

This paper is structured as follows. Related Works reviews the existing works that are related to the proposed technique. In Proposed Residual Deformable Split Depth-Wise Separable U-Net Method, proposed network's architecture is explained. Following that, Residual Deformable Encoder presents and discusses the experimental outcomes. Finally, Middle Processing Module provides a conclusion.

Related Works

To solve the issue of precisely defining the boundary of liver imaging in CT scans, Xie et al. [22] developed a method based on an adaptive dynamic residual network (ADR-net). Initially, a data augmentation method is employed to rise the number of training samples and the robustness of the network. This method prevents precision loss by introducing a pooling component. This enables the CNN to acquire more precise features for various pooling domains at various iteration rates, enabling the DA-pooling layer to extract feature values that most closely match the original input image. Additionally, residual structures are used to prevent gradient explosion and accuracy degradation. First, after convolution, the batch normalisation (BN) operation is implemented to increase training accuracy and hasten model convergence. This operation solves the issue of the vanishing gradient. Second, Relu activation functions are replaced by PRelu activation functions. This non-linear activation function, along with batch normalisation-based DA-pooling and residual networks, enables the network to obtain and combine high- and low-level data more effectively. Conditional random fields (CRF) are employed to improve the margins of the liver area, preventing the segmentation results from over- or under-segmentation; however, this can lead to a loss of spatial information.

Aghamohammadi et al. [23] introduced a two-path feature extraction model based on CNN (TPFE-CNN) that uses 3 input images instead of only one to identify the boundary of the liver and tumours. This method initially applies a Z-score normalisation technique to the input images in order to produce a more distinct liver boundary. More details on the difference between the liver and the tumour can be found in this normalised image. Then, a new encoding approach called local direction of gradient was used to encode images in order to extract more important characteristics. Even in areas close to connecting organs, the provided encoding image is very good at identifying the liver's border. Then, using the input image and two additional generated images as input data, a cascading CNN structure is used to extract semi-global and local features. Instead of utilising a complicated deep CNN model with numerous hyperparameters, a simple but efficient model is used to reduce the testing and training times.

Kushnure and Talbar [24] introduced the MS-UNet with a feature recalculation method to segment the liver and LT. To represent multi-scale information and broaden the CNN's receptive field (RF), the bottleneck Res2Net module is used. It characterises the input's global and local data by extracting multi-scale characteristics. As a result, the input feature is collected hierarchically to represent the precise information of the input feature, which is represented with numerous features of various sizes. Additionally, a squeeze and excitation (SE) network is used to channel-wise recalibrate the multi-scale characteristics. It enhances the network's capacity for learning and focuses on the object's more notable properties. The Res2Net module and SE network combined improve the network's ability to represent features and its learning potential. The Res2Net module reduces the parameters and computational complexity of the proposed network.

In order to segment the LT on CT images with the best possible generalisation and accuracy, Chung et al. [25] developed an automatic auto-context CENet (AutoCE-Net). Using a liver-prior branch, the auto-context method is applied to a neural network. In order to calculate the foreground likelihood of liver, the liver-prior branch is heavily supervised. The last auto-context layers then combine the previous with deep contexts. To further define the shape of a liver, one more branch that is similarly deeply supervised is introduced along with the auto-context framework. Sparse contours are trained using a self-supervised approach that serves as implicit contour attention. By penalising the ground truth contour picture depending on confidence in its final forecast, the network achieves self-supervision. Thus, the network pays close attention to its mistakes rather than utilising the whole ground truth contour or self-attention.

Chi et al. [26] introduced a hybrid dense X-net to concurrently extract liver and LT features in the same CT slice. Here, an additional de-convolution section is introduced to the DenseU-Net for the prediction of the liver region, and the feature pyramid maps of the estimated liver are utilised as additional features for the LT segmentation. In order to effectively exploit intra- and inter-slice contextual characteristics, the conventional 3D U-Net is improved. Slices in the CT volume's spatial coherence are taken into account, and the 3D de-convolutional and convolutional filters are simplified in order to forecast the liver area and tumours. In order to reflect the differences between the predicted outcomes and the actual results with regard to the region contours, a shape loss function based on the active contour theory was also implemented. Additionally, the conventional DICE loss function is also modified to better take into account the

negative data when there is no liver or tumour. The network parameters are then successfully optimised using the joint loss function. The entire network is trained using the multi-task learning approach, allowing for more effective optimisation of the liver segmentation and tumour segmentation branches.

Meng et al. [27] introduced a multi-scale feature fusion system called TransFusionNet. It is a multi-scale data fusion system with a multi-layer local feature extraction module, a transformer-based semantically feature extraction module, and a multi-scale fusion decoder that can learn spatial and semantic information features. By combining the global context of input images, the model can successfully identify and segment liver tumours. In addition, it successfully extracts the edge features of the target objects based on the edge extraction module. With the help of the transfer learning technique, the network is able to combine the features from the three datasets, which greatly reduces overfitting and boosts segmentation accuracy in the final trained model. Finally, compressed distillation is used to deploy the model to the embedded microprocessor, enabling the model to segment medical images in real-time.

Zhou et al. [28] present a contour-aware feature merged network (CFNet) to enhance and optimise liver segmentation on CT images, taking into account the slice correlation continuity and the lack of edge contour information in the fuzzy edge of the liver. To analyse contextual information, CFNet uses convolutional bi-directional long short-term memory (CBiLSTM). Additionally, a CBiLSTM-AG module is created by fusing the local features of an attention gate (AG) with the global contextual information of CBiLSTM. CBiLSTM high-dimensional information is fused using the AG to remove superfluous features. Additionally, it is suggested to use the Shape-Net to expand the shape pattern of the liver using latent space data, which lessens interference from fuzzy edges.

With adaptive region growing (ARG) and graph cuts (ARG + GC), Yang et al. [29] designed a suitable semi-automatic method for segmenting LTs. First, an ARG with a manually chosen seed for every tumour area is used to obtain the initial segmentation findings for LTs and the region of interest (ROI) containing the tumours. The ROIs are then improved via non-linear mapping with Gaussian fitting based on the intensity distributions of the initially segmented tumour regions. The tumours from the ROIs are then precisely and successfully extracted using graph cuts that incorporate the improved and gradient information. The technique is noise-insensitive and does not need liver pre-segmentation or a time-consuming, laborious training process. It considerably lessens the complexity of segmentation and the computational burden. However, this model requires a high computational cost for high-resolution images.

To solve the difficulties of fixed geometrical convolutional kernels being mismatched with liver and LTs with irregular shapes, as well as the loss of spatial information of input images caused by strided convolutional and pooling processes, Lei et al. [30] proposed a deformable encoder-decoder network (DefED-Net). An atrous ladder spatial-pyramid pooling (ALSPP) is designed using a multi-scale dilation rate and is applied to learn context information better than pyramid pooling (PP) for segmentation. The capacity of DefED-Net to represent features is improved by the use of deformable convolution, which enables the model to learn convolution filters with adaptable spatial data. However, this model increases the training time because of the increased number of parameters. Table 1 shows the comparison of the existing liver and LT segmentation.

Proposed Residual Deformable Split Depth-Wise Separable U-Net Method

The RDSDSU-Net uses the deformable convolution layer (DCL) with a residual structure to produce FMs. FMs generated from these learned sample sites have a stronger capacity to generalise complicated structures than the standard convolution approach. Also, DP is used to downsample the FMs. More structural properties of LTs are extracted using this pooling. In the middle processing module, a convolutional spatial and channel features split graph network (CSCFSGnet) is introduced. To improve the expression capability of LTs features, convolution is split to get spatial and channel features. Then, global contextual LT information in spatial and channel features is obtained by a graph convolution network (GCN). The decoder part consists of SPC instead of an upsampling layer. Compared to other approaches, this method speeds up network convergence and yields high-resolution output. The architecture of the proposed RDSDSU-Net is shown in Fig. 1.

Residual Deformable Encoder

Although several improved U-Nets have been available for segmenting medical images, their liver and LT segmentation accuracy is only moderate. Here are two factors that affect U-Net: First, U-Nets disregard the shape information of an image's objects by using convolutional kernels with predetermined geometrical shapes. Second, using pooling and strided convolution results in the loss of detailed information about the spatial environment. So, the proposed RDSDSU-Net uses the deformable convolution (DC) layer (DCL) [31] with the residual structure to produce FMs to avoid information loss due to a predetermined geometrical structure. To increase the DC kernel's adaptability to various shapes and positions of the liver and LT, learnable offsets are introduced to every sample position of the RF. The DCL determines the precise sample

Table 1 Comparison of the existing liver and LT segmentation

Author	Technique	Application	Dataset	Advantages	Limitations	Metrics					
						DICE	VOE	RVD	ASD	RMSD	MSD
Xie et al. [22]	ADR-net	Liver segmentation	3DIRCADb	It enhanced the convergence speed and generalisation capability	Spatial information loss occurs due to the presence of large number of pooling	✓	✓	✓	✓	-	✓
Aghamohammadi et al. [23]	TPFE-CNN	Liver and LT segmentation	Non-public dataset from Milad Hospital in Iran	It does not require any extra parameters other than the source images to create the segmented areas	Only consider local features and ignore global features	-	✓	✓	✓	-	✓
Kushnure and Talbar [24]	MS-UNet	Liver and LT segmentation	3DIRCADb	Avoid spatial information loss	The concatenation of various scale features in MS-UNet necessitates a large amount of memory	✓	✓	✓	✓	-	✓
Chung et al. [25]	AutoCENet	Liver segmentation	MICCAI-SLiver07, 3Dircadb, CHAOS	It concentrates on accuracy and the generalisation ability	Only slices with clear liver or LT borders produce accurate segmentation findings	✓	-	-	✓	-	-
Chi et al. [26]	X-net	Liver and LT segmentation	LiTS and 3DIRCADb	It extracts intra-slice characteristics that can highlight additional disparities between liver tissues, LT, and other areas	Difficult to capture the various shapes of the liver and LT due to the usage of fixed kernel sizes	✓	✓	✓	✓	✓	✓
Meng et al. [27]	TransFusionNet	Liver tumour and vessel segmentation	LiTS and 3DIRCADb	Multi-scale fusion avoids spatial information loss	Large amounts of memory and parameters are needed due to the multi-scale data fusion,	✓	✓	-	-	-	-
Zhou L et al. [28]	CFNet	Liver segmentation	LiTS, CHAOS, and 3DIRCADb	Precisely segment the liver and LT by extracting both global and local features	It is difficult to capture the various shapes of the liver and LT due to the usage of fixed kernel sizes	✓	✓	✓	✓	✓	✓
Yang et al. [29]	ARG + GC	LT segmentation	3DIRCADb	It doesn't need a time-consuming, laborious training process and liver pre-segmentation	Accuracy is low	✓	✓	✓	✓	✓	✓
Lei et al. [30]	DefED-Net	Liver and LT segmentation	LiTS and 3DIRCADb	Avoid loss of spatial contextual data by using DC	Avoid channel characteristics and just capture spatial context information	✓	✓	✓	✓	✓	✓

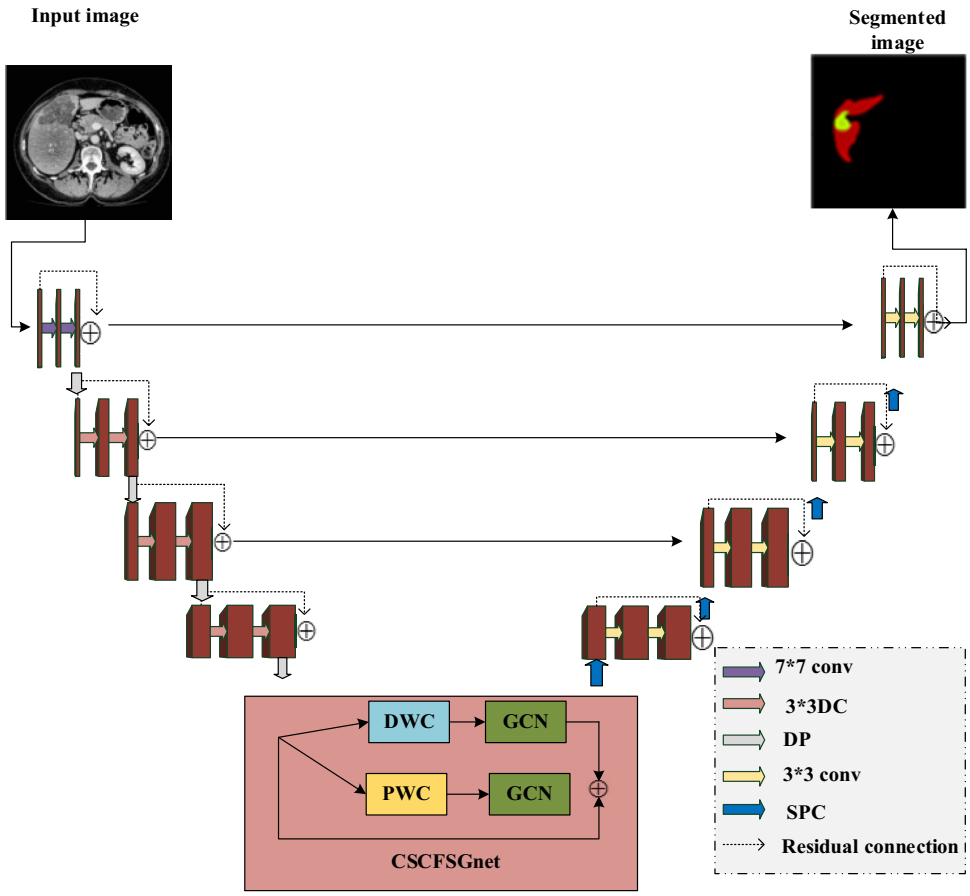

Fig. 1 Architecture of proposed RDSDSU-Net

sites based on the liver and LT's shapes. Figure 2 illustrates the flowchart for the DCL. First, the offset field is formed to obtain the offset values. Next, bilinear interpolation is applied to obtain the deformable feature pictures.

Assume that the output and input FMs are $\tilde{z}$ and y, respectively. Then, the output of the DCL is described in Eq. (1).

$$\tilde{z}(f_o) = \sum_{f_n \in M} w(f_n) \times y(f_o + f_n + \Delta f_n) \tag{1}$$

where Δf_n denotes the offsets added to each sampling location, f_o denotes a pixel's location in sampling field, M represents sampling field, f_n denotes each regular squared sample

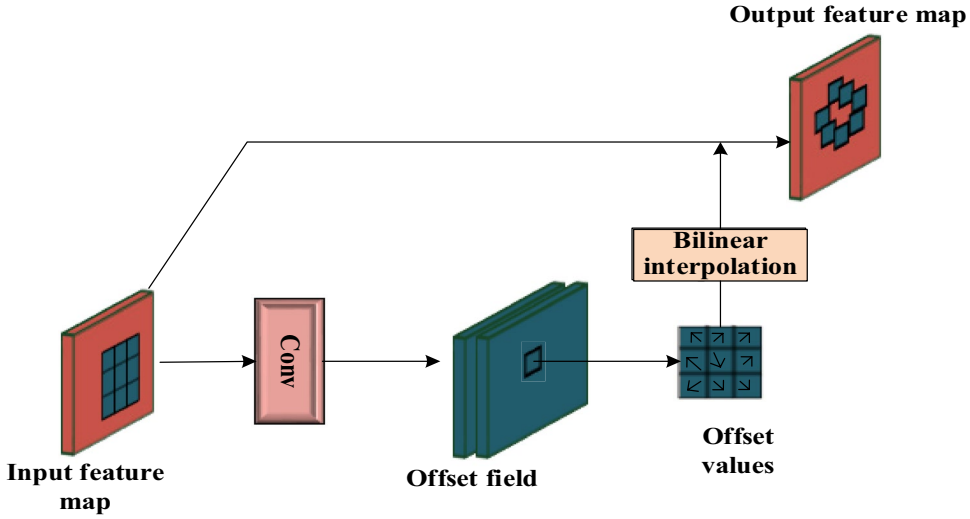

Fig. 2 Architecture of DCL

location, and $w(f_n)$ denotes the associated weight values. The offset is generated by applying a convolution layer (CL). Also, CL used to calculate the offset must have the same dilation rate and spatial resolution as the CL used to obtain the features from the offset FM. When CL receives an input FM with P channels from the DCL, the associated offset map has $2P$ channels since each channel contains 2 offset maps in the horizontal and vertical directions.

The bilinear interpolation is utilised to execute the pixel value selection of the final sampling location since the offset Δf_n is often a float number and the DC's sampling location becomes irregular. The last sampling position's pixel value $y(f)$ is defined in Eq. (2).

$$y(f) = C(x_i,\ r_j) \tag{2}$$

where r_j represents the 4 adjacent pixels used at the irregular sampling point computation, w_i represents the associated weight, and $C(.)$ represents the bilinear interpolation kernel. $C(.)$ is represented in Eq. (3).

$$C(w_i,\ r_j) = w_1\, r_1 + w_2\, r_2 + w_3\, r_3 + w_4\, r_4 \tag{3}$$

Back propagation of Eqs. (2) and (3) is used to learn the offset values during training. Finally, a new FM is produced after pixel values for every sampled location have been determined. Large convolutional kernels are preferable to small kernels for extracting coarse liver regions while doing liver segmentation tasks. But, a smaller kernel is more beneficial for capturing contour information. As a result, large kernels are used in the first DCL, and smaller ones are used in succeeding layers.

DP is used instead of max pooling to downsample the input FMs the DC to avoid the loss of spatial information. The DP process for the $\left(k^{th}, l^{th}\right)$ patch, given a pooled FM z in size of P, is represented in Eq. (4).

$$z(k,l) = \frac{\sum_{f \in bin(k,l)} (f_o + f + \Delta f_{kl})}{P^2} \tag{4}$$

$$\left\{\Delta f_{kl} | 0 \le k, j < P\right\}$$

where Δf_{kl} denotes the offsets added to the usual squared pooling positions and f_o denotes the top-left corner of the liver and LT region. Over the input feature pictures, a fully connected layer calculates the offset fields. Δf_{kl} values are learned by back propagation to create DP FMs. The DP fully extracts the liver and LT's structural characteristics compared to other pooling layers with fixed sampling locations.

Middle Processing Module

Previous research [10] demonstrates the efficiency of extracting contextual characteristics from pictures using a deep convolution network with atrous convolutional blocks and multi-kernel branching. However, utilising these blocks causes the model to become more complex. The U-Net used in medical image segmentation depends only on the concatenation of multi-scale characteristics in a straightforward manner. It does not fully exploit channel and spatial information. At this time, many U-Nets give importance to fully exploiting spatial information but avoid channel information, which affects segmentation accuracy. Therefore, CSCFS-Gnet is introduced to enhance segmentation performance and capture spatial and channel context information. The flowchart for CSCFSGnet is shown in Fig. 3.

To improve the capability of liver and LT features to express themselves, the first convolution is split to extract

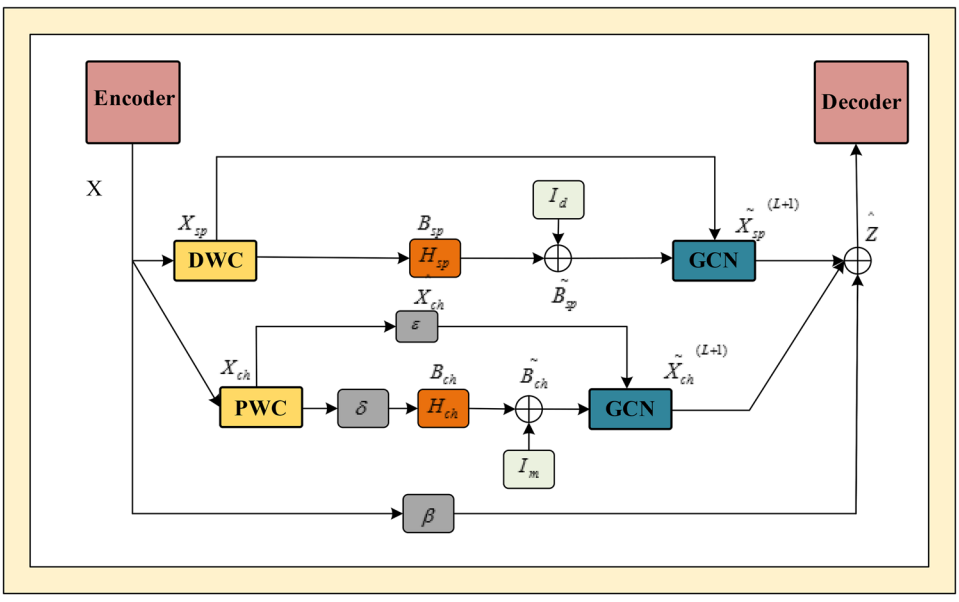

Fig. 3 Flowchart for CSCFS-Gnet

spatial and channel features. Then, to collect global contextual liver and LT features in spatial and channel features, a GCN is used. The feature graph's adjacency matrix (AM) is created using the sobel gradient operator.

The proposed CSCFSGnet model divides the depth wise separable convolution [32] to acquire spatial and channel characteristics. Features ($h * w * d$) from the encoder are given to CSCFSGnet. In the DWC stage, d filters with $m * m * 1$ size are utilised to individually convolute every channel to obtain spatial features X_{sp}. In the PWC stage, features are convolved to get channel features X_{ch} using d filters with $1 * 1 * d$ size.

The proposed module uses a gradient operator to build an AM. Because it is more logical to establish edges between nodes when node values in X_{sp} and X_{ch} vary rapidly since these nodes provide more information, thus, to identify these changes, a gradient operator was used. It transforms X_{sp} into the adjacency spatial matrix $B_{sp_{\frac{m\times m}{r^4}}} \in \Re^{d\times d}$, while X_{ch} to adjacency channel matrix $B_{ch} \in \Re^{\frac{}{r^4}}$, where r represents the downsample rate.

The AM is created by multiplying the outcomes of H_x and H_y's convolution in the X and Y directions, respectively. The AM is created by multiplying these results, which is expressed in Eqs. (5) and (6).

$$\tilde{B}_{sp} = H_x(X_{sp})H_y(X_{sp}) + I_d \tag{5}$$

$$\tilde{B}_{ch} = H_x(\delta(X_{ch}))H_y(\delta(X_{ch})) + I_m \tag{6}$$

where I_d and I_m represent the identity matrices. In Eq. (6), operator δ changes the channels' locations inside the grid and downsamples the features of the channels such that the sobel operator can process the channel dimensions while also minimising the number of parameters.

A fully connected graph was created once the AM had been produced. Afterwards, by creating an AM, GCNs [33] expand the concept of convolution to graph-structured data. GCNs were first used in knowledge graphs, and more recently, they have been used to extract features from images. GCNs can increase the RF and collect contextual data. Contrary to ordinary convolution, graph convolution generalises the neighbourhood definition and enables long-range sharing of information in a single layer, making it a useful and efficient module. The downsampling operator ε is applied for X_{ch} in order to be compatible with $B_{ch\ \text{and}}$ is given in Eq. (7).

$$X^{\wedge}{}_{ch} = \varepsilon(X^{\wedge}{}_{ch}) \tag{7}$$

$$\tilde{X}_{sp}^{(L+1)} = a(\tilde{E}_{sp}^{-\frac{1}{2}} \tilde{B}_{sp} \tilde{E}_{sp}^{-\frac{1}{2}} (X_{sp}^{T})^{(L)} w_{sp}) \tag{8}$$

$$\tilde{X}_{ch}^{(L+1)} = a(\tilde{E}_{ch}^{-\frac{1}{2}} \tilde{B}_{ch} \tilde{E}_{ch}^{-\frac{1}{2}} (\hat{X}_{ch}^{T})^{(L)} w_{ch}) \tag{9}$$

where w_{sp} and w_{ch} are trainable weight matrix, $\tilde{E}_{ch}$ and $\tilde{E}_{sp}$ represent degree matrix of graph, and α represents non-linear activation function.

Three layers of graph convolution are used for the proposed CSCFSGnet to compute the representation of channel and spatial information separately. The proposed CSCFSGnet combines two separate graphs to learn improved feature representation after generating a graph structure. The final finely tuned feature is calculated as $\hat{Z} = \tilde{X}_{sp}^{(3)} + \tilde{X}_{ch}^{(3)} + \beta(X)$. β is a 3×3 convolution operation and + represents PW summation.

Decoder

Bilinear upsampling or a transposed convolutional layer is commonly employed in the original U-Net to recover the downsampled features to the final segmentation result. However, the segmentation performance would be reduced by these algorithms' tendency to produce the chequerboard artefacts effect, which displays discontinuous or inconsistent predictions across neighbouring picture patches. Therefore, SPCs have been added to the proposed RDSDSU-Net's decoder to mitigate this impact. Here, the low-resolution CSCFSGnet FM is the sub-pixel convolution's input. The intermediate FMs with channels R^2 are created using a sequence of convolution operations, where R is the original FM's enlargement factor. The expanded high-resolution FMs, which are R times bigger than the input, are then obtained by periodic shuffling. In this method, the SPC layer reduces the chequerboard artefact issues.

Furthermore, due to the addition of several convolutions, the decoder now has a greater capacity for non-linear transformation than it had when using just bilinear upsampling, which may also boost the segmentation performance. Also, instead of doing a simple copy concatenation as U-Net does, residual deformable encoder features are concatenated to the decoder through summation. The benefit of utilising summation is that it results in fewer parameters in the subsequent layers since the number of FMs is not increased.

Loss Function

The proposed RDSDSU-Net approach is required to train to predict whether each pixel is in the background or foreground. One of the most commonly used loss functions is the cross-entropy, which is described in Eq. (10).

$$l_{cross\ entropy} = -(R\log(\hat{R}) + (1-R)\log(1-R^{\wedge})) \tag{10}$$

However, the tumour usually only takes up a tiny portion of the image. For these tasks, the cross entropy loss is not

Table 2 Layer settings for the proposed RDSDSU-Net

	Stage	Layer	Kernel size
Encoder	1	DCL	7*7
		DCL	7*7
		BN	
		DP	3*3
	2	DCL	3*3
		DCL	3*3
		BN	
		DP	3*3
	3	DCL	3*3
		DCL	3*3
		BN	
		DP	3*3
	4	DCL	3*3
		DCL	3*3
		BN	
		DP	3*3
Middle processing module	1	DWC	3*3*1
		PWC	1*1*d
	2	3GCL	3*3
Decoder	1	SPC	
		Conv	3*3
		Conv	3*3
	2	SPC	
		Conv	3*3
		Conv	3*3
	3	SPC	
		Conv	3*3
		Conv	3*3

suitable. It is important to note that the DICEloss works well with uneven samples. This metric effectively calculates the amount of segmentation outcome overlap with the related ground truth. The DICEloss is described in Eq. (11).

$$l_{dice} = 1 - \frac{2 < R, \hat{R} >}{((\|R\|_1 + \left\| \hat{R} \right\|_1))} \tag{11}$$

where $R \in (0, 1)$ and $0 \leq \hat{R} \leq 1$ are true. The label and the segmentation outcomes are represented by R and $\hat{R}$, respectively. $< R, \hat{R} >$ represents dot product.

However, using the DICEloss easily impacts the back propagation and makes training challenging. The two losses are thus combined to define the loss function.

$$l_{loss} = l_{crossentropy} + l_{dice} \tag{12}$$

Table 2 shows the proposed RDSDSU-Net's entire design, including the number of network stages and convolution filter size.

The whole pipeline for training and testing is shown in Fig. 4. The input CT dataset is initially pre-processed, after which the RDSDSU-Net is trained for the liver and LT, tested on test data, and network performance is assessed using some metrics.

Results and Analysis

The proposed RDSDSU-Net is applied to the 3DIRCADb and LiTS to assess the performance of RDSDSU-Net in the liver and LT segmentation by comparing it with the existing approaches. The details of the two datasets, evaluation metrics and parameter setup, are presented in the following sections. The outcomes of the ablation study are then discussed, along with comparisons to other models.

Dataset

The performance evaluation of the proposed RDSDSU-Net will be conducted in Python using the 3DIRCADb dataset

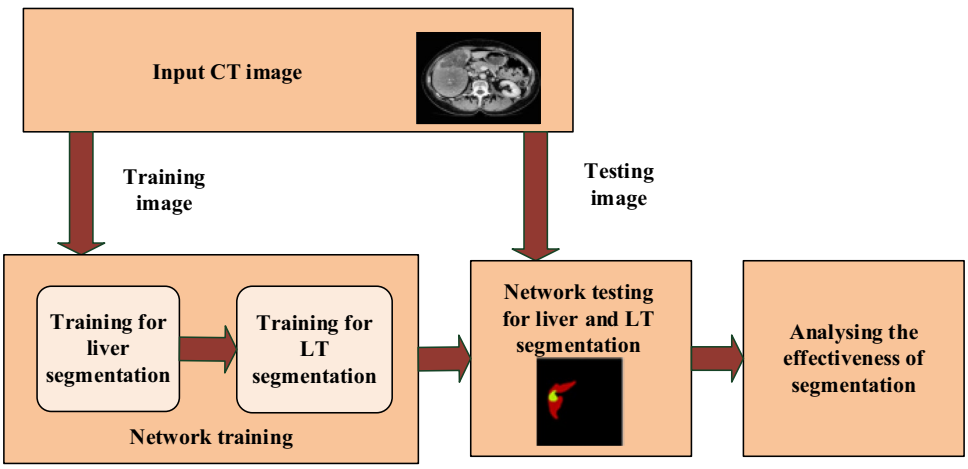

Fig. 4 An illustration of the RDSDSU-Net pipeline for liver and LT segmentation

[34] and LiTS dataset [35]. The LiTS dataset comprises 201-3D abdominal CT scans in the axial plane, of which 131 scans are utilised for training and 70 scans are used for testing. In the 3DIRCADb dataset, the data from 15 patients (2295 total images) were utilised for training, and the data from the remaining five patients (525 total images) were utilised for testing [17]. The image intensity of all scans is reduced by [− 200, 250] HU and applied normalisation to them to eliminate interferences and improve liver areas. In RDSDSU-Net, the liver and LT segmentation are carried out independently and separately.

Experimental Setup and Metrics for Evaluation

The proposed RDSDSU-Net is implemented using Python on a system equipped with Windows 10 operating system, an Intel i7-10510U CPU (16 GB memory), and NVIDIA GeForce MX250 GPU. The parameters used for the proposed RDSDSU-Net are shown in Table 3. The performance of the proposed RDSDSU-Net was evaluated by comparing it with existing methods using per-case metrics such as DICEsimilarity coefficient (DSC), relative volume difference (RVD), volumetric overlap error (VOE), root mean square symmetric surface distance (RMSD), and average symmetric surface distance (ASD).

RVD, VOE, and DICE overlap measures concentrate on the inside of the segmentation target. DICE is calculated using Eq. (13).

$$DICE = \frac{2|G \cap S|}{|G| + |S|} \tag{13}$$

where G represents ground truth (label) and S represents segmented outcome.

RVD is calculated using Eq. (14).

$$RVD = \frac{|G| - |S|}{|S|} \tag{14}$$

VOE is calculated using Eq. (15).

$$VOE = \left(1 - \frac{|G \cap S|}{|G \cup S|}\right) \tag{15}$$

The surface distance between segmentation outcomes and labels is measured using both ASD and RMSD. The ASD calculates the average distance, while the RMSD computes the maximum distance. ASD and RMSD focus more on the segmentation target's similarity in shape.

Equation (16) represents the nearer distance from a random voxel rv to $R(G)$.

$$D(rv, R(G)) = \min_{R_G \in R(G)} \left\| rv - R_G \right\| \tag{16}$$

RMSD and ASD are calculated using Eqs. (18) and (17), respectively.

$$ASD(G, S) = \frac{1}{|R(G)| + |R(S)|}\left(\sum_{R_G \in R(G)} D(R_G, R(S)) + \sum_{R_S \in R(S)} D(R_S, R(G))\right) \tag{17}$$

$$RMSD(G, S) = \sqrt{\frac{1}{|R(G)| + |R(S)|}} \times \sqrt{\sum_{R_G \in R(G)} D^2(R_G, R(S)) + \sum_{R_S \in R(S)} D^2(R_S, R(G))} \tag{18}$$

where $R(G)$ represent sthe number of surface voxels of G. *DICE*, *VOE*, and *RAVD* are measured in %, and *ASD* is measured in millimetres (mm). For DICE, 100% represents the best segmentation, 0% represents the poorest segmentation, and for RAVD and VOE, the ideal segmentation is 0%, whereas the poor segmentation is 100%. These measurements are in percentage form. The optimum segmentation for ASD is 0 mm, while the highest value exhibits the poorest segmentation. The higher value has no limit.

Ablation Study

The proposed RDSDSU-Net enhances context information in the liver and tumour segmentation by using three key contributions: a residual deformable encoder instead of an ordinary encoder, the integration of GCN, and the usage of SPC instead of a bilinear upsampling layer. To demonstrate its effectiveness, extensive tests are conducted on both LiTS and 3DIRCADb datasets. The results show that the proposed RDSDSU-Net outperforms previous methods on both datasets.

Effectiveness of Residual Deformable Encoder

The networks considered in ablation study are:

a U-Net
b U-Net + DC
c U-Net + residual DC
d U-Net + residual DC + DP

Table 3 Parameter setting for proposed RDSDSU-Net

Parameters	Values
Initial learning rate (L_r)	0.001
Decay learning rate	$L_r = L_r \times \left(1 - \frac{number\ \ of\ iterations}{total\ \ iterations}\right)^{0.9}$
Learning rate for offset convolutional layers in DC	$L_{r2} = L_r \times 0.01$
Optimiser	Adam

The experimental outcomes of these models are shown in Tables 4 and 5. Compared to U-Net + residual DC and U-Net + DC, U-Net captures many regions other than the liver. Both DC and RDC assist U-Net in directing attention to the important areas and eliminating the unnecessary background areas, while residual DC achieves faster network convergence and provides more precise edge predictions. However, without the downsampling layer, it increases the number of parameters. Tables 4 and 5 show that using DC improves U-Net's segmentation accuracy. The residual architecture increases segmentation accuracy while simultaneously increasing U-Net convergence. The DP fully extracts the structural characteristics and reduces the training time by reducing the number of parameters. Thus, using residual DC and DP in U-Net achieves better segmentation results than the other three models.

Effectiveness of CSCFSGnet

Three models are used in this ablation study to assess the effectiveness of using the CSCFSGnet in liver and LT segmentation. The three models used in this ablation study are:

a U-Net + atrous spatial pyramid pooling [ASPP]
b U-Net + ALSPP
c U-Net + CSCFSGnet

The experimental findings in Tables 4 and 5 demonstrate that ASPP and ALSPP assist U-Net in increasing the accuracy of liver and LT segmentation. However, U-Net + CSCFSGnet has high segmentation accuracy because it extracts global context information. Compared to U-Net models, the outcomes from U-Net + CSCFSGnet are more accurate. Additionally, U-Net + ALSPP and U-Net + ASPP only enhance the model's representational capabilities in terms of capturing spatial context data. In comparison to U-Net + ALSPP and U-Net + ASPP, U-Net + CSCFSGnet has a high DICE score. Thus, CSCFSGnet improves the accuracy of segmentation by capturing both spatial and channel features.

Table 4 Ablation study comparison on LiTS dataset

Models	DICE (%) for liver	DICE (%) for LT
U-Net	93.89 ± 1.19	82.19 ± 6.27
U-Net + DC	95.20 ± 1.18	84.57 ± 6.22
U-Net + residual DC	95.8 ± 1.13	85.6 ± 6.17
U-Net + residual DC with DP	96.36 ± 1.12	91.36 ± 3.14
U-Net + ASPP	93.35 ± 1.14	83.25 ± 5.14
U-Net + ALSPP	94.23 ± 1.11	84.21 ± 5.18
U-Net + CSCFSGnet	96.23 ± 1.07	86.14 ± 3.08
U-Net + bilateral up sampling	91.22 ± 1.17	81.26 ± 5.18
U-Net + residual connection + SPC	95.22 ± 1.12	90.07 ± 3.12
Proposed RDSDSU-Net	97.35 ± 0.02	90.15 ± 0.01

Table 5 Ablation study comparison on 3DIRCADb dataset

Models	DICE (%) for liver	DICE (%) for LT
U-Net	91.78 ± 1.17	51.19 ± 6.26
U-Net + DC	93.20 ± 1.16	54.57 ± 6.23
U-Net + residual DC	95.6 ± 1.11	52.6 ± 6.18
U-Net + residual DC with DP	96.34 ± 1.10	82.37 ± 3.15
U-Net + ASPP	91.34 ± 1.12	62.27 ± 5.15
U-Net + ALSPP	93.21 ± 1.09	63.20 ± 5.19
U-Net + CSCFSGnet	97.24 ± 1.05	87.15 ± 3.10
U-Net + bilateral up sampling	90.20 ± 1.15	60.27 ± 5.19
U-Net + residual connection + SPC	96.20 ± 1.10	81.08 ± 3.13
Proposed RDSDSU-Net	98.21 ± 0.04	93.25 ± 0.04

Effectiveness of Residual SPC Decoder

The models used to evaluate the effectiveness of residual SPC decoder are:

a U-Net + bilateral upsampling
b U-Net + residual connection + summation + SPC

In the normal U-Net model, transposed convolution and bilateral upsampling are used in the segmentation. U-Net + bilateral upsampling layer recovers the downsampled features, but this model creates a chequerboard artefacts effect, affecting the segmentation output. As shown in Tables 4 and 5, U-Net + residual connection + summation + SPC creates accurate segmentation results by avoiding chequerboard artefact effects.

According to Tables 4 and 5, it is evident that the DICE score of LT segmentation on the 3DIRCADb dataset is better than that of the LiTS dataset. This could be attributed to several factors. Firstly, the 3DIRCADb dataset contains a smaller number of images compared to the LiTS dataset. As the proposed method performs better on smaller datasets due to having fewer parameters to learn, it is less likely to overfit the data. Secondly, the 3DIRCADb dataset has more complex LTs than the LiTS dataset. The proposed method is better suited for handling the complex tumours in the 3DIRCADb dataset due to its use of advanced techniques such as the convolutional spatial and channel features split graph network. This method helps to focus on the relevant features in the images by giving equal importance to both spatial and channel features.

Table 6 Quantitative liver segmentation results on the 3DIRCADb dataset using various methodologies

Techniques	DICE (%)	RVD (%)	RMSD (mm)	VOE (%)	ASD (mm)
X-net [26]	96.68 ± 0.41	0.29 ± 0.23	2.95 ± 2.74	6.45 ± 2.08	3.07 ± 0.48
MS-UNet [24]	96.13 ± 1.02	0.25 ± 0.14	-	5.57 ± 2.10	4.08
DefED-Net [30]	96.60 ± 1.08	0.23 ± 0.11	12.76 ± 3.43	5.65 ± 2.81	2.61 ± 0.84
Proposed RDSDSU-Net	98.21 ± 0.04	0.12 ± 0.10	1.46 ± 2.34	4.24 ± 2. 24	2.05 ± 0.43

Table 7 Quantitative liver segmentation results on the LiTS dataset using various methodologies

Techniques	DICE (%)	RVD (%)	RMSD (mm)	VOE (%)	ASD (mm)
X-net [26]	96.8 ± 0.05	-	-	-	-
CFNet [28]	97.4	− 4.963	-	4.963	-
DefED-Net [30]	96.30 ± 1.01	1.46 ± 0.12	70.05 ± 3.10	6.88 ± 2.10	1.37 ± 0.23
Proposed RDSDSU-Net	97.35 ± 0.02	0.92 ± 0.11	1.06 ± 1.2	14.12 ± 2.76	1.14 ± 0.12

Table 8 Quantitative LT segmentation results on the 3DIRCADb dataset using various methodologies

Techniques	DICE (%)	RVD (%)	RMSD (mm)	VOE (%)	ASD (mm)
ARG + GC [29]	85 ± 5	− 0.05 ± 0.12	16.1 ± 0.37	26.03 ± 0.07	68.45 ± 0.36
X-net [26]	69.11 ± 5.83	− 0.68 ± 0.33	23.6 ± 2.12	36.09 ± 12.61	15.1 ± 0.61
MS-UNet [24]	84.15 ± 2.34	0.22 ± 0.55	-	27.36 ± 11.43	16.4 ± 0.42
DefED-Net [30]	66.25 ± 6.62	0.81 ± 0.20	70.05 ± 3.10	34.28 ± 13.43	11.21 ± 0.63
Proposed RDSDSU-Net	93.25 ± 0.04	− 0.07 ± 0.10	11.06 ± 0.02	14.12 ± 2.76	10.21 ± 0.33

Table 9 Quantitative LT segmentation results on the LiTS dataset using various methodologies

Techniques	DICE (%)	RVD (%)	RMSD (mm)	VOE (%)	ASD (mm)
TransFusionNet [27]	91.0	-	-	− 0.018	-
X-net [26]	76.4 ± 0.03	-	-	-	-
DefED-Net [30]	87.52 ± 5.32	0.52 ± 0.10	64.25 ± 4.87	23.85 ± 14.62	17.41 ± 0.28
Proposed RDSDSU-Net	90.15 ± 0.01	0.43 ± 0.09	32.06 ± 2.42	14.16 ± 3.62	12.12 ± 0.13

Comparison with State-of-the-Art Technique

Existing models that are used for segmenting the liver and LTs like ARG + GC [29], TransFusionNet [27], CFNet [28], X-net [26], MS-UNet [24], and DefED-Net [30] are taken into consideration as comparison techniques to demonstrate the effectiveness of the proposed RDSDSU-Net. In comparison to other techniques, Tables 6, 7, 8, and 9 illustrate that the proposed RDSDSU-Net offers the best quantitative results (ASD, DICE, RMSD, RVD, and VOE). The segmentation results from various methodologies are shown in Fig. 5. The proposed method accurately segments the small LT, unlike existing methods, by giving equal importance to both channel and spatial features. As a result, the RDSDSU-Net outcomes in the fifth column give precise segmentations of the liver and LTs than the existing networks. Additionally, the RDSDSU-Net concentrates on the liver area while avoiding the effect of other organs, resulting in smoother segmentation borders than other techniques. Figure 5 demonstrates that the RDSDSU-Net accomplishes accurate spatial and channel feature extraction, which is beneficial for enhancing the liver and LT segmentation.

Performance Comparison: Operations, Model Size, and Training Parameters

In addition, we calculate the computational costs and number of training parameters for each network, as presented in Table 10. The proposed RDSDSU-Net method has the lowest number of operations (186.23 GFLOPS), smallest model size (32.56 MB), and fewest training parameters (10,532,286) compared to other existing approaches such as X-net, TransFusionNet, CFNet, and DefED-Net. This means that the proposed RDSDSU-Net method achieves low inference and training time as well as a less complex model structure, which could potentially result in

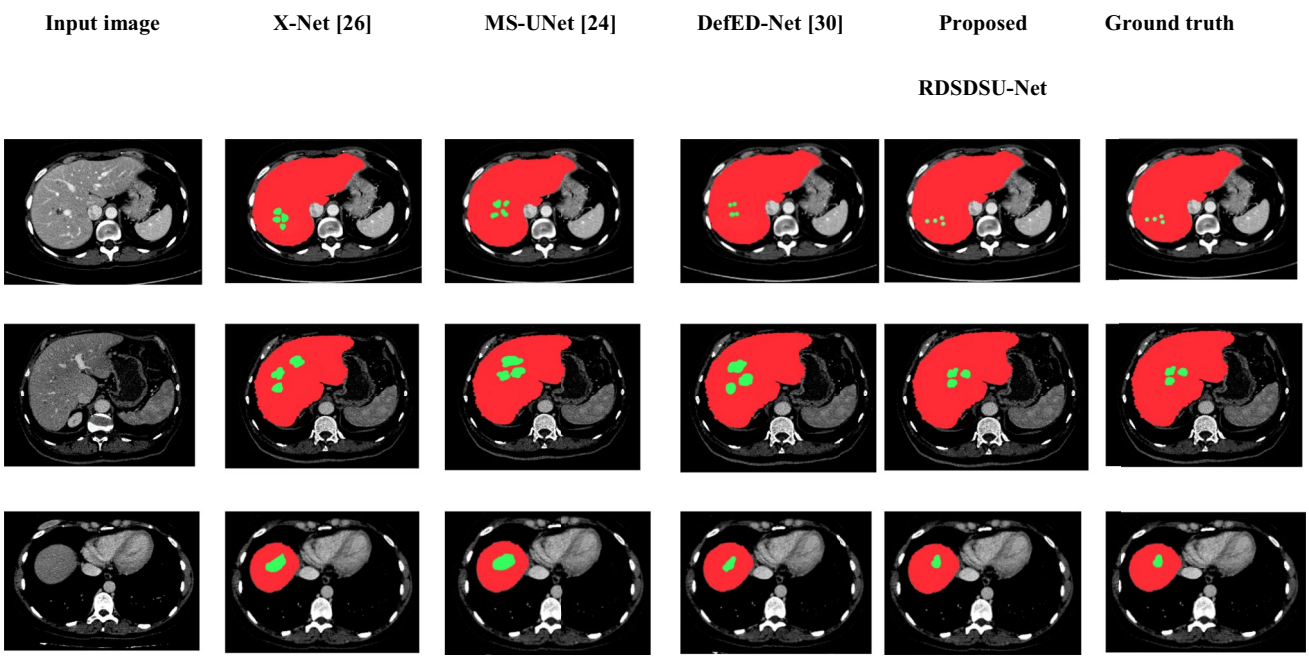

Fig. 5 Results of segmenting the liver and LTs using various methods

improved performance in real-world applications. The proposed method achieves the lowest number of operations, the smallest model size, and the fewest training parameters by utilising a few key techniques. First, the method employs a novel architecture design that optimizes the network's depth and width while reducing the number of parameters. This design relies on a depth-wise separable convolutional layer that separates the spatial and channel-wise convolutional operations, resulting in a significant reduction in the number of parameters and computational complexity. Also, the residual deformable encoder features are not immediately concatenated by the decoder in the proposed technique. Instead, it combines the features using a summation process, which reduces the number of parameters without increasing the quantity of feature maps.

Table 10 Performance comparison of various network architectures

Method	Operations (GFLOPS)	Model size (MB)	Training parameters
X-net [26]	402.13	22.34	73,818,112
TransFusionNet [27]	374.09	20.87	64,736,579
CFNet [28]	515.13	147.18	74,072,814
DefED-Net [30]	222.44	59.6	14,529,959
Proposed RDSDSU-Net	186.23	32.56	10,532,286

Conclusion

The segmentation of small and irregularly shaped tumours has caused difficulties for the robustness of the segmentation model because of the variations in liver and LT forms and sizes. In this situation, RDSDSU-Net is suggested to enhance segmentation results. To get better irregular-shaped liver and LT features, residual DCL with pooling is introduced into U-Nets. APP and ALSPP also result in the loss of detailed information, even though both are good at improving context information. So the CSCFSGnet has been proposed to acquire spatial and channel properties and to better capture global contextual information. Finally, SPC is used in place of the bilinear upsampling layers to eliminate the chequerboard artefact in the segmentation. Also, the training parameters are proportionally decreased by using summation in combining encoder and decoder features. This method demonstrates that spatial and channel information use is more crucial than feature fusion in segmenting the liver and LT. Experiments on the 3DIRCADb and LiTS datasets show that the proposed RDSDSU-Net improves liver and LT segmentation accuracy. In future work, the segmentation performance of RDSDSU-Net is enhanced by using post-processing techniques like conditional random field [36] and level set [37].

Author Contribution All the authors have participated in writing the manuscript and have revised the final version. All authors read and approved the final manuscript.

Data Availability All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Saumiya S, S. Wilfred Franklin. The first draft of the manuscript was written by Saumiya S and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.Conceptualization: Saumiya S; Methodology: Saumiya S, S. Wilfred Franklin; Formal analysis and investigation: Saumiya S, S. Wilfred Franklin; Writing - original draft preparation: Saumiya S; Writing - review and editing: Saumiya S, S. Wilfred Franklin; Supervision: S. Wilfred Franklin.

Declarations

Ethical Approval This article does not contain any studies with human participants and/or animals performed by any of the authors.

Informed Consent There is no informed consent for this study.

Conflict of Interest The authors declare no competing interests.

References

1. Calderaro J, Ziol M, Paradis V, Zucman-Rossi J (2019) Molecular and histological correlations in liver cancer. Journal of hepatology. 71(3):616-30.
2. Anwanwan D, Singh SK, Singh S, Saikam V, Singh R (2020) Challenges in liver cancer and possible treatment approaches. Biochimica et Biophysica Acta (BBA)-Reviews on Cancer. 1873(1):188314.
3. Ranjbarzadeh R, Saadi SB (2020) Automated liver and tumor segmentation based on concave and convex points using fuzzy c-means and mean shift clustering. Measurement. 150:107086.
4. Sivanandan R, Jayakumari J (2021) Ultrasound liver tumour active contour segmentation with initialization using adaptive Otsu based thresholding. Research on Biomedical Engineering. 37(2):251-62.
5. Dorgham OM, Alweshah M, Ryalat MH, Alshaer J, Khader M, Alkhalaileh S (2021) Monarch butterfly optimization algorithm for computed tomography image segmentation. Multimedia Tools and Applications. 80(20):30057-90.
6. Zhao J, Li D, Xiao X, Accorsi F, Marshall H, Cossetto T, Kim D, McCarthy D, Dawson C, Knezevic S, Chen B (2021) United adversarial learning for liver tumor segmentation and detection of multi-modality non-contrast MRI. Medical Image Analysis. 73:102154.
7. Liu T, Liu J, Ma Y, He J, Han J, Ding X, Chen CT (2021) Spatial feature fusion convolutional network for liver and liver tumor segmentation from CT images. Medical Physics. 48(1):264-72.
8. Gao Q, Almekkawy M (2021) ASU-Net++: A nested U-Net with adaptive feature extractions for liver tumor segrementation. Computers in Biology and Medicine. 136:104688.
9. Aboelenein NM, Songhao P, Koubaa A, Noor A, Afifi A (2020) HTTU-Net: Hybrid Two Track U-Net for automatic brain tumor segmentation. IEEE Access. 8:101406-15.
10. Hong Y, Mao XW, Hui QL, Ouyang XP, Peng ZY, Kong DX (2021) Automatic liver and tumor segmentation based on deep learning and globally optimized refinement. Applied Mathematics-A Journal of Chinese Universities. 36(2):304-16.
11. Tran ST, Cheng CH, Liu DG (2020) A multiple layer U-Net, U n-Net, for liver and liver tumor segmentation in CT. IEEE Access. 9:3752-64.
12. Zhang D, Chen B, Chong J, Li S (2021) Weakly-supervised teacher-student network for liver tumor segmentation from non-enhanced images. Medical Image Analysis. 70:102005.
13. Tang W, Zou D, Yang S, Shi J, Dan J, Song G (2020) A two-stage approach for automatic liver segmentation with Faster R-CNN and DeepLab. Neural Computing and Applications. 32(11):6769-78.
14. Liu Z, Han K, Wang Z, Zhang J, Song Y, Yao X, Yuan D, Sheng VS (2021) Automatic liver segmentation from abdominal CT volumes using improved convolution neural networks. Multimedia Systems. 27(1):111-24.
15. Dong X, Zhou Y, Wang L, Peng J, Lou Y, Fan Y (2020) Liver cancer detection using hybridized fully convolutional neural network based on deep learning framework. IEEE Access. 8:129889-98.
16. Kaur A, Kaur L, Singh A (2021) GA-UNet: UNet-based framework for segmentation of 2D and 3D medical images applicable on heterogeneous datasets. Neural Computing and Applications. 33(21):14991-5025.
17. 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 transactions on medical imaging. 39(5):1316-25.
18. Jiang H, Shi T, Bai Z, Huang L (2019) Ahcnet: An application of attention mechanism and hybrid connection for liver tumor segmentation in ct volumes. Ieee Access. 7:24898-909.
19. Bai Z, Jiang H, Li S, Yao YD (2019) Liver tumor segmentation based on multi-scale candidate generation and fractal residual network. IEEE Access. 7:82122-33.
20. Anter AM, Hassenian AE (2019) CT liver tumor segmentation hybrid approach using neutrosophic sets, fast fuzzy c-means and

adaptive watershed algorithm. Artificial intelligence in medicine. 97:105-17.
21. Liu Z, Song YQ, Sheng VS, Wang L, Jiang R, Zhang X, Yuan D (2019) Liver CT sequence segmentation based with improved U-Net and graph cut. Expert Systems with Applications. 126:54-63.
22. Xie X, Zhang W, Wang H, Li L, Feng Z, Wang Z, Wang Z, Pan X (2021) Dynamic adaptive residual network for liver CT image segmentation. Computers & Electrical Engineering. 91:107024.
23. Aghamohammadi A, Ranjbarzadeh R, Naiemi F, Mogharrebi M, Dorosti S, Bendechache M (2021) TPCNN: two-path convolutional neural network for tumor and liver segmentation in CT images using a novel encoding approach. Expert Systems with Applications. 183:115406.
24. Kushnure DT, Talbar SN (2021) MS-UNet: A multi-scale UNet with feature recalibration approach for automatic liver and tumor segmentation in CT images. Computerized Medical Imaging and Graphics. 89:101885.
25. Chung M, Lee J, Park S, Lee CE, Lee J, Shin YG (2021) Liver segmentation in abdominal CT images via auto-context neural network and self-supervised contour attention. Artificial Intelligence in Medicine. 113:102023.
26. Chi J, Han X, Wu C, Wang H, Ji P (2021) X-Net: Multi-branch UNet-like network for liver and tumor segmentation from 3D abdominal CT scans. Neurocomputing. 459:81-96.
27. Meng X, Zhang X, Wang G, Zhang Y, Shi X, Dai H, Wang Z, Wang X (2021) Exploiting full Resolution Feature Context for Liver Tumor and Vessel Segmentation via Fusion Encoder: Application to Liver Tumor and Vessel 3D reconstruction. arXiv preprint arXiv:2111.13299. Accessed on: August 25, 2022.
28. Zhou L, Deng X, Li W, Zheng S, Lei B (2021) A contour-aware feature-merged network for liver segmentation based on shape prior knowledge. Neurocomputing. 457:389-99.
29. Yang Z, Zhao YQ, Liao M, Di SH, Zeng YZ (2021) Semi-automatic liver tumor segmentation with adaptive region growing and graph cuts. Biomedical Signal Processing and Control. 68:102670.
30. Lei T, Wang R, Zhang Y, Wan Y, Liu C, Nandi AK (2021) DefED-Net: Deformable encoder-decoder network for liver and liver tumor segmentation. IEEE Transactions on Radiation and Plasma Medical Sciences. 6(1):68-78.
31. Dai J, Qi H, Xiong Y, Li Y, Zhang G, Hu H, Wei Y (2017) Deformable convolutional networks. InProceedings of the IEEE international conference on computer vision (pp. 764–773).
32. Chollet F (2017) Xception: Deep learning with depthwise separable convolutions. InProceedings of the IEEE conference on computer vision and pattern recognition (pp. 1251–1258).
33. Kipf TN, Welling M (2016) Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907. Accessed on: August 25, 2022.
34. https://www.kaggle.com/datasets/nguyenhoainam27/3dircadb. Accessed on: August 27, 2022.
35. https://www.kaggle.com/datasets/andrewmvd/liver-tumor-segmentation
36. Liu Y, Yao J, Lu X, Xie R, Li L (2019) DeepCrack: A deep hierarchical feature learning architecture for crack segmentation. Neurocomputing. 338:139-53.
37. Le TH, Quach KG, Luu K, Duong CN, Savvides M (2018) Reformulating level sets as deep recurrent neural network approach to semantic segmentation. IEEE Transactions on Image Processing. 27(5):2393-407.