Def-UNet with Feature Fusion and Recalibration for Liver Segmentation in Multi-Modality CT Images

Bindu Madhavi A. Tummala and Soubhagya Sankar B. Barpanda *

School of Computer Science and Engineering, VIT-AP University, near Vijayawada, Andhra Pradesh, India
Email: bindumadhavi.t@vitap.ac.in (B.M.A.T.); soubhagya.barpanda@vitap.ac.in (S.S.B.B.)
*Corresponding author

Abstract*—In this paper, we have proposed a deformable encoder-decoder neural network for liver segmentation from multi-modality Computed Tomography (CT) images. Liver segmentation is a predominant step to taking conclusive action toward liver disease detection, therapeutic decision planning, and post-operation assessment. The computed tomography scan has become the default choice of medical practitioners to determine hepatic anomalies. However, due to improvements in image acquisition protocols, imaging data is growing making the manual delineation process burdensome and tedious for clinicians and becoming reliant on expert proficiency and experience. Furthermore, automatic liver segmentation is challenging due to complicated anatomy, shape variance, and less contrast variation within itself and its tumors, between its neighboring organs like the heart, and spleen, and even discontinuity in liver contours. Moreover, normal convolutions with fixed feature patterns cannot predict irregular liver patterns Thus, our proposed Def-UNet for liver segmentation is developed by modifying the encoder convolution method by deformable convolutions and skip connections by local feature recalibration which sends high-level feature information to the decoder side. The deformable convolution is computationally less expensive and best suited for shape-variant medical images. Further, the adaptive recalibration through a Squeeze-and-Refine network helps to learn the channel-wise interdependencies and gather the salient details from the fusion applied high-level features. As a bridge module, we have employed an atrous pyramid pooling module to capture the spatial information from the low-level features with the help of dissimilar receptive fields. These methods help the Def-UNet to enhance the accuracy and greatly reduce the computational burden of the other DL-based segmentation methods. The efficacy of the proposed method is experimented on two datasets Combined (CT-MRI) Healthy Abdominal Organ Segmentation (CHAOS) and 3DIRCADb that are publicly available. The experimental result analysis illustrates that the proposed model has attained a dice similarity coefficient of 0.966 and 0.972 for liver segmentation.

***Keywords*—liver segmentation, deformable convolution, deep learning, squeeze and refine networks, encoder-decoder architecture**

I. Introduction

Medical imaging became the de-facto standard in clinical diagnosis due to its functionality of representing internal organs and tissues visually through 2D or 3D images, called slices. The rapid growth and availability of different imaging modalities like Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) made the availability of medical data abundant and brought the need for computerized medical image analysis. The preliminary step in medical image analysis is image segmentation. This process delineates the organ of interest from the background modality image and thus helps the clinician to have a detailed organ interpretation that leads to disease prediction followed by treatment. Accurate and efficient medical image segmentation assists doctors by reducing the time-consuming process of inspecting a larger number of slices to identify specific organs and lesions [1]. Among the human organs, the liver is highly introspected among many death-causing diseases. The liver is the largest organ present on the right side of the abdomen and is responsible for various biological processes like blood regulation, and toxic breakdowns. It is the heaviest gland weighing approximately 2% of the total human body weight. Liver diseases like cancer and cirrhosis are dangerous to human health and the occurrence and mortality rate are sixth and fourth among the disease-related deaths in the world. Liver disease detection is done through biological imaging tests. Among them, CT is the widely used method due to its high spatial resolution, non-invasive, and low cost. Liver segmentation is important for applications like accurate surgical planning in liver resection and liver transplantations that need the assessment of the preoperative liver size of the donor, portal vein embolization in major hepatectomy, and even in printing 3D models. The liver is one of the most complex organs to segment because of its variable shape and densities which are caused by diverse pathologies like fat, fibrosis, iron deposits, and tumors [2].

Manuscript received April 16, 2024; revised July 4, 2024; accepted August 2, 2024; published February 14, 2025.

doi: 10.18178/joig.13.1.64-72

The challenges present in liver imaging that lower the performance outcomes are shown in Fig. 1. First, the CT patterns of the liver are very much like its adjacent organs such as lungs, kidneys, and muscles making the delineation process a tedious task. Secondly, the anatomy of the liver is different from patient to patient due to factors like age, lifestyle habits, and more particularly the formation of fat. Thirdly, the differences between the scanners cause variations in the shape and location of the liver. In routine medical practices, the doctors manually segment the liver and it is time-consuming and overburdened. It can also lead to a poor assessment because of human errors and the decision is dependent on the experience of the expert. Moreover, the advancements in medical imaging brought an increase in volume per patient scan and became the main concern for doctors to perform manual delineation. So, several semi-automatic segmentation methods are developed based on image processing techniques like thresholding and region-growing [3]. These methods take advantage of selecting seed point and threshold set value manually and the results are user-dependent and show extreme inter and intra-observer differences. The next popular approach is the Active Contour Model (ACM) which performs the segmentation by contour delineation [4]. This approach attempts to draw the contour of the target object by using several optimization techniques [5]. The limitation of this approach is the large variance shown in intensity distribution due to the unclear boundaries effect. Contrarily, the shape-based models were developed to overcome the difficulty and as expected they outperformed the intensity-based methods but were limited due to the little availability of liver shapes database and the differences in shape and size from healthy to diseased subjects. Because of all the aforementioned issues, automated liver segmentation is still a challenging and interesting field of research for many researchers.

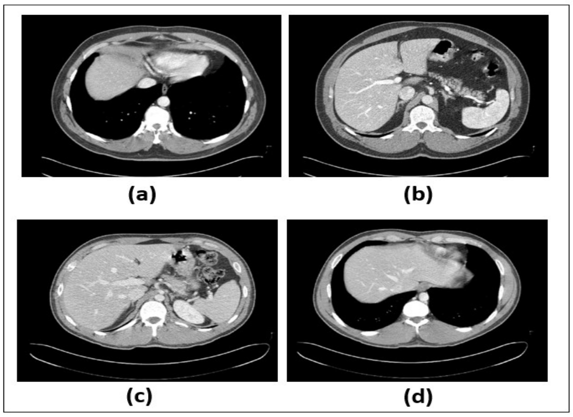

Fig. 1. Challenges in liver segmentation: (a) merged boundaries (b) disconnected liver components (c) less enhanced liver tissues (d)irregular shapes at different ranges.

II. Literature Review

Recent advancements and state-of-the-art performances in the computer vision field were brought by the Deep Learning (DL) concept that came into practice due to the availability of heavy computational power. The most used deep learning model is the Convolutional Neural Network (CNN) due to its capacity for feature self-learning and the competency to learn from a large set of images through a set of layers [6]. This accelerated the researchers to implement DL in sophisticated medical image analysis steps like organ and tumor segmentation, disease prediction, and classification and help the medical clinicians in regular diagnosis which in turn improves the service quality of the patients [7]. The Fully Convolutional Neural Network (FCN) built on the concept of the encoder-decoder deep model was the most used architecture for liver segmentation. The stepping stone for this decision is the UNet architecture designed for microscopic image segmentation [8]. A two-step integrated encoder-decoder architecture was proposed to perform liver tumor segmentation [9]. The classic improvement is by using pretrained networks via transfer learning with a frozen encoder. Almost all UNet-inspired architectures for liver segmentation performed the exact preprocessing techniques in the precise order of HU windowing, contrast enhancement, and normalization. Furthermore, many post-processing techniques like Conditional Random Field (CRF), level cut, graph cut, thresholding, and a few random forest methods are used to improve the delineation accuracy [10, 11–14]. Later, many add-ons to UNet are introduced to perform liver segmentation. Using curriculum learning liver and its tumors are extracted jointly with a UNet-based model [15]. UNet++[16] proposed U-Nets with different depths to avoid improper fusion of different levels of information. The DenseUNet [12] fuses the 2D and 3D feature representations to exploit the inter and intra-slice variations. An optimized convolutional neural network is introduced and concentrated on parameter count reduction [17]. A 3D attention module is added to the UNet with residual learning to improve the feature learning [18]. A volumetric attention module integrated with Mask-RCNN is used to achieve liver and tumor segmentation [19]. V-Net [20] also achieved 3D segmentation by using 3D convolutions with residuals. The 3D architectures yield better accuracy by exploring large spatial dimensions but fail at the extreme usage of computational power, memory, and even the number of network parameters. The residual connections and the attention modules help the network to concentrate more on the ROI and enhance its convergence speed. A modified UNet is developed to perform segmentation by altering the skip connections with the residual paths [21]. All these modifications on UNet simply lead to an increase in encoder stages and eventually the performance degrades due to the contextual information loss. However, the additions in skip connections help to some extent to hold the high-resolution contextual information but the spatial loss cannot be recovered completely. In segmentation, the anatomical semantics of the organ plays an important role in the performance and this information is lost in encoder-decoder architectures. To meet this impotency, a fusion of multi-scale feature extraction was introduced and it led to the development of segmentation performance. The multiscale features are evaluated with the help of dilated convolutions in the encoder-decoder network [22]. Densely connected convolutional residual connections are used along with the residual skip connections to gather the rich contextual information. Dilated convolutions with different rates and different pooling kernels are used to extract the multiscale features [23–25]. Thus, multiscale CNN architectures have shown a substantial advancement in liver segmentation. However, for the end-users of these models, the clinicians must work with both healthy and unhealthy subjects drawn from different single-modality scanners along with the contrast injection. All the developed models restrict themselves from this level of generalizability. They considered datasets only from unhealthy patients. So, to handle this issue we have proposed a novel Def-UNet architecture using feature reconstruction, multi-scale feature fusion, and deformable convolutions to improve the liver segmentation performance from the multi-phase CT images. The overall workflow of the proposed model is shown in Fig. 2.

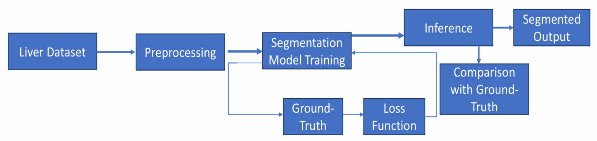

Fig. 2. Overall workflow of the proposed model.

Our contributions are:

(1) Feature extraction in Def-UNet is done by a residual deformable convolution module that extracts the features without the loss of spatial information at higher layers.
(2) The images are fed to the network on multiple scales to ensure the fusion of features at all levels of spatial information. The training image set consists of healthy subjects from three different CT scanners and unhealthy subjects from a single CT scanner.
(3) The adaptive feature recalibration at skip connections layers improves the learning and segmentation accuracy of the network.
(4) The pyramid pooling module is used at the bottleneck of the architecture with different dilation rates for obtaining better contextual information.
(5) The Def-UNet achieves better accuracy than the other approaches used in the study along with the benefit of less memory and computation requirements due to the use of dilation and deformable convolutions.

The remainder of the paper is organized as follows: Section III explains the details of the proposed methodology and its advantages. Section IV discusses the training infrastructure of the network and Section V discusses the effectiveness of the method through segmentation results and comparative result analysis. Section VI follows the conclusion.

III. Materials and Methods

In our work, a deformable UNet (Def-UNet) is proposed to perform liver segmentation from multi-phase CT images. The architecture is shown in Fig. 3. It is evident from the diagram that the proposed method is a U-shaped architecture with four modules: encoder, decoder, skip connections, and, the bottleneck that connects the encoder-decoder streams. The encoder is responsible for high-level feature learning and these features carry more critical spatial information that helps to carry the delineation process more effectively. And, the encoder-extracted features are transferred to the respective decoder layer via skip connections to help the decoder reconstruct the image. It clearly shows that the quality of the encoder directly influences the overall result. So, to have better feature map representations within the encoder residual deformable convolutions are used in the place of normal 2D convolutions. The input to the encoder is fed at multiple scales that help in extracting more useful contextual information from all the encoder levels. The up-gradation done in the skip connection path is the I-block. This block applies the fusion of low and high-level feature maps and the features are recalibrated to present rich spatial information by applying depth-wise SENet [21]. Now, these improvised features present only the important characteristics of the decoder and help it rebuild the organ within its shape, size, and location. The bottleneck layer is implemented by parallelly connected pooling kernels with different dilation rates to give a better upsampling performance to the decoder. All the modules are explained below clearly stating the importance and advantage of using them at that point within the proposed architecture.

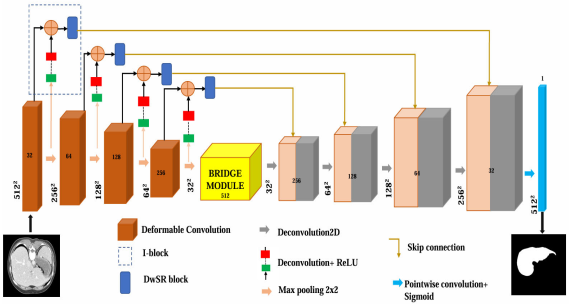

Fig. 3. Def-UNet architecture.

A. Deformable Encoder

In medical segmentation, the organ shape and contour play a major role in the success of the segmentation accuracy. However, the traditional convolutional kernels use a fixed geometrical shape for all the inputs. They use a rectangular grid with fixed filter sizes like 3×3, 5×5, and so on. Each time the grid is moved with a fixed offset to generate the feature maps. Therefore, the traditional or regular convolutions do not consider the shape information provided by the input image or a set of images. Meanwhile, the liver has an invariable shape and size, and even location based on the phase of the CT scan or due to the applied augmentations. Therefore, we have employed deformable convolution at the encoder side which can be defined in simple terms as learnable atrous convolution. This deformable grid information best suits the irregularly shaped liver and improves feature learning. This type of convolution adaptively selects the receptive field based on the scale of the image and thus correctly learns the fine contour information. The learnable offset of deformable convolution made it understand different geometrical shapes and orientations more easily. Fig. 4 demonstrates the impact of a normal convolution and a deformable convolution on liver segmentation. In general, the normal convolution comprises (convolution + batch normalization activation layer) as a single unit. In deformable convolution, an extra fourth layer is added called offset. This layer is responsible for adaptive dilation based on the scale of the image. The equation is given as follows: Let p and q denote the input and output feature maps respectively. Let R be the regular grid used for convolution operation. Then the output feature map q is derived as:

$$q(a_0) = \text{X}\ _{w}(a_n) \times p\ (a_0 + a_n) \quad (1)$$

where, $a_n \in R$, w represents the weight a_0 represents the pixel location and a_n represents all the adjacent pixels within R. The feature map q derived from deformable convolution is:

$$q(a_o) = \text{X}\ _{w}(a_n) \times p(a_0 + a_n + \Delta a_n) \quad (2)$$

where, $a_n \in R$, Δa_n is the variable offset value, a fractional number that is responsible for the irregular grid. The fractional value cannot be added to the samplings, so the bilinear interpolation method is used to calculate the pixel values of the deformed position through:

$$p(a) = \text{X}_G(b,a).p(b) \quad (3)$$

where, *G(b, a)* is two-dimensional and represented as:

$$G(b,a) = g(b_i,a_i).g(b_j.a_j) \quad (4)$$

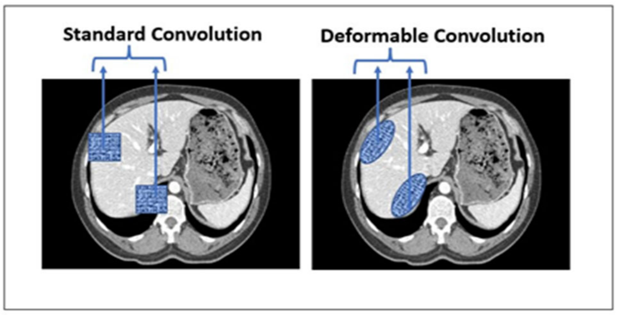

Fig. 4. Normal convolution vs deformable convolution.

The offset is calculated by applying convolution to the input feature map (refer to Fig. 5.). In deformable convolutions, the dilation rate of the kernel should align with the requirements of the current layer. The feature map with N channels obtains offset maps of 2N channels, where each set of N channels represents offset in different directions individually. These offsets are learned during training. The process of training involves bilinear interpolation as described by Eqs. (3) and (4), which allows the network to adjust sampling positions dynamically for improved feature extraction. To capture the liver shape precisely a large convolutional kernel 5 ×5 along with a residual connection is used to improve the convergence speed of the network and remove vanishing gradients that occur due to large receptive fields.

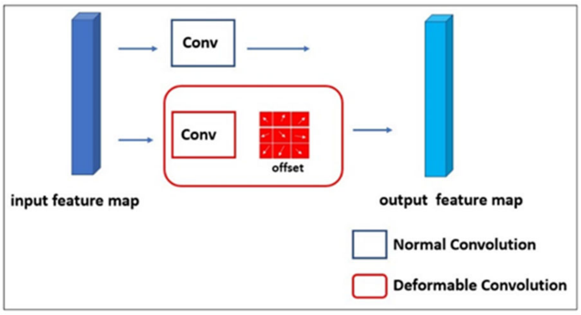

Fig. 5. Feature structure for normal and deformable convolution.

B. I-Block

The rich contextual information that helps the decoder reconstruct the segmented image comes from the upsampling layers through skip connection pathways in encoder-decoder-style neural architectures. The improvement block (I-block) is present at the skip pathway that joins the encoder and decoder stages layer-wise.

This block improves the feature representations and helps the decoder with high-level rich contextual information. Let us consider S stages where i = 1, 2, 3—S-1. At every i^th^ stage, deformable convolution (convolution+batch normalization + activation + offset) is performed to obtain the output feature map set where i = 1, 2, 3—S-1. At every ith stage, deformable convolution (convolution+batch normalization + activation + offset) is performed to obtain the output feature map set of$_i$.

$$of_i = (f_1, f_2, -- f_k) \quad (5)$$

where k denotes the number of feature channels. The output feature set f$_i$ of the ith stage is reconstructed through max-pooling P$_i$ followed by upsampling U$_i$ (deconvolution + ReLU) and now the input features if$_i$ of the ith stage are fused as shown in Eq. 6. This feature fusion FF$_i$ at every stage improves the quality of the feature set that represents the crucial spatial information.

$$FF_i = if_i + U_i(P_i(of_i)) \quad (6)$$

The improvised features are now given to the Depth-wise Squeeze and Refine (DwSR) module to recalibrate the features. The feature recalibration is generally done by Squeeze-Excitation (SE) block [21] or the Convolutional Block Attention (CBA) module [22] that improves the needed features and refines the useless features. However, the use of maximum global or average pooling operation to reduce the parameters causes information loss. One method loses more information and the other fails in identifying the important contributions of the features. In the concept of learning dependency within the channels, these blocks altered the independence between them. Instead, we have used depth-wise separable convolution in our DwSR module to perform feature recalibration. Depth-wise convolution applies 1×1 point convolutions on a single feature map, thereby reducing computationally hungry parameters without compromising the effect of normal convolution. Given a set of feature maps F to the DwSR block generates the recalibrated features F as output. The DwSR block performs two operations: depth-wise squeeze and refine. In depth-wise squeeze operation, two depth-wise convolutions (dwconv) are used with the prescribed kernel size for the first dwconv and global setting for the second dwconv. The early layers have large-sized feature maps so we have used both specific and global kernel settings and moving further in layers the size of the feature is reduced so only global kernel setting is used. To retrieve complete information from the features, the activation function is not used between the two depth-wise convolutions. The refined operation refines the calibrated features by removing the channels with zero values because they have a negative impact on the network. To achieve this, a sigmoid function is used:

$$F_{sigmoid} = (DW_2(DW_1(F))) \qquad (7)$$

where, DW$_1$ and DW$_2$ are the weights of two depth-wise convolutions. A scaling factor s is added to achieve the global properties of the feature.

$$f_c = s_c.f_c \qquad (8)$$

where $F = \{F_1, F_2, F_3, ..., F_c\}$ and . denotes the multiplication of the feature map and the scaling factor s_c per channel and these values change accordingly with the input feature set size. The feature fusion followed by recalibration improves the quality of the feature representations and thus concatenates with the respective decoder layers to increase the network performance. This entire improvisation is represented as I-block in the network diagram depicted in Fig. 6.

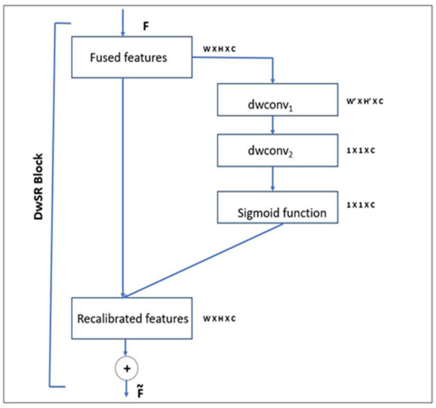

Fig. 6. The structure of I-Block.

C. Bridge Module Using Pyramid Pooling

The output of the last encoder stage has its feature size low and thus the data transferring to the decoder layers become less contextual. In organ segmentation, contextual information plays a crucial role in defining the accuracy of the segmented object. The last encoder layer has a very low volume and low-level features to have a good start for the decoder, pyramid pooling is done on the bridge module. The bridge module establishes the connection between the last layer of the encoder to the first layer of the decoder. To increase the quality of the features, a pooling pyramid with dilation variants is applied as shown in Fig. 7. In this module, we have used depth-wise separable atrous convolutions to reduce the complexity of the architecture without compromising the quality of the extracted features.

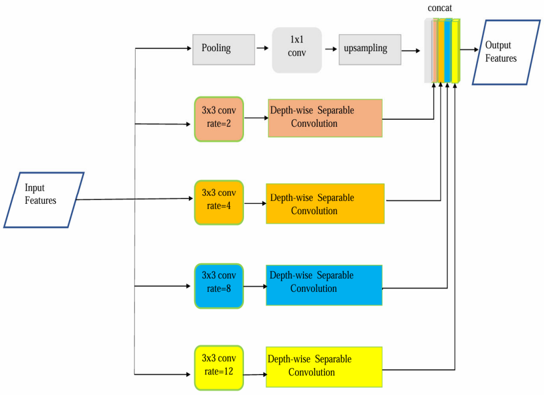
concat
Pooling
1x1 conv
upsampling
Output Features
Input Features
3x3 conv rate=2
Depth-wise Separable Convolution
3x3 conv rate=4
Depth-wise Separable Convolution
3x3 conv rate=8
Depth-wise Separable Convolution
3x3 conv rate=12
Depth-wise Separable Convolution

Fig. 7. The pyramid pooling architecture.

The depth-wise separable convolution uses depth-wise convolution ensued by 1×1 convolution to preserve the depth and the equation is:

$$F(i) = {}^{\mathrm{X}}f[i + a.k]w[k] \qquad (9)$$

where, a denotes the atrous rate of the convolution used and it is responsible for the field of view. The different dilation rates used obtain different size variant features. The atrous rates used are a = 3, 6, 9 and these multi-scale features are combined with recalibrated features and offered to the decoder layer for segmentation. The proposed Def-UNet thus uses both the high and low levels and modifies them in terms of quality through feature recalibration and pyramid pooling methods to improve the accuracy of the segmentation architecture. The individual feature reconstruction at each skip connection helps to bind the more precise semantic information that is very helpful for the decoder in upsampling the segmented image. To make the low-level features semantically rich, the bridge module performs parallel atrous depth-wise convolutions at different rates to extract low-level multi-scale features. The depth-wise convolutions not only reduce the parameters but also help the network to concentrate more on learning rather than on memorizing the parameters. The improvement block (I-block) combines the high-level features and modified high-level features obtained through fusion with reconstructed features and recalibrated through the squeeze-refine network. The symbols represent fusion and concatenation respectively. The batch normalization helps in smoothing the gradient. The selected data sets are preprocessed and then the network learns these images through training using the backpropagation method finally the performance is evaluated through the evaluation metrics

IV. Result and Discussion

This section discusses the parameters that are needed to train the network. Training is a procedure that makes the network learn from the input feed and perform the intended task.

A. Datasets and Preprocessing

To train and test the proposed network we have used 3DIRCADb (3D Image Reconstruction for Comparison of Algorithm Database) [26] and CHAOS (Combined Healthy Abdomen Organ Segmentation) [27] publicly available datasets. The 3DIRCADb dataset contains 20 CT scans acquired from 10 men and 10 women in the portal venous phase with various CT scanners from European hospitals. Among 20 scans, 15 scans contain 75% of tumors. The CHAOS dataset also contains 20 CT scans drawn from three different CT modalities. All these scans are from 20 different healthy liver donors without any lesions or tumors. Complete information about the datasets is listed in Table I. Both the datasets have several challenges for liver segmentation as shown in Fig. 8. These challenges are also listed as follows:

(1) Atypical or irregular liver shapes.

(2) Blurred boundaries with adjacent or neighboring organs such as the stomach, spleen, heart, and pancreas due to similar Hounsfield range.

(3) The liver region is represented with different Hounsfield ranges due to contrast injection.

(4) Artifacts in CT scans like metal artifacts, beam hardening, scatter, and metal artifacts bring noise to the images.

(5) Different shapes of the liver within the patients due to its variant anatomy.

All these challenges are addressed by our proposed system along with the good performance on segmentation accuracy. Before feeding the data to training, the data has to be preprocessed. Data preprocessing steps are important because they improve the quality of the CT slices by clearing the CT value artifacts through simple steps: HU clipping, contrast enhancement, and normalization. Every organ in the human body has some CT value or number called a Hounsfield unit within the range of −1,000 to 1,000. The liver region falls within −40 to 50 HU. The HU clipping step sets the HU window range for all the CT slices between −250 HU to 200 HU which brings visualclarity to the liver region. A contrast enhancement and noise amplification method called Contrast Limited Adaptive Histogram Equalization (CLAHE) [28] is used to improve the brightness of the liver region so that the target can become clearer. To improve the convergence speed of the network all the input images are normalized to the range [0, 1]. All these steps of data preprocessing change the raw images into a useful form to feed the neural network (Fig. 9). Generally, deep learning network models need more data to learn and perform well. The less amount of data leads to overfitting of the model and yields poor performance. However, image acquisition and annotation is a time burden in the medical field because the gathering of medical images along with the ground truths should be done only by clinical experts. Data secrecy also plays a major role in the public availability of medical data sets. The data augmentation technique solves this problem by enlarging the small data sets by augmenting the original images. Data augmentation is a part of preprocessing where the original images are duplicated by applying some transformations. The augmenting methods used in this paper are transpose, rotation 90, horizontal and vertical flips. So for one original image set (ct slice along with its ground truth), four duplicate image sets are generated which in turn will expand the dataset. The sample augmented set is shown in Fig. 8. The totals of images in the datasets before and after augmentation and complete information about the datasets are displayed in Table I. Thus the augmentation procedure progresses the training of the Def-UNet by resolving the overfitting problem and improving the generalizable capability of the network.

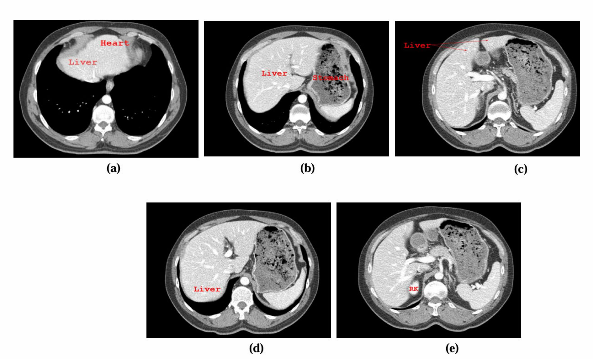

Fig. 8. Challenges present in two datasets: (a) attached boundaries with neighboring organ heart (b) contrast-enhanced liver tissues (c) disconnected liver components (d) less enhanced liver tissues (e) partial volume effect with right kidney.

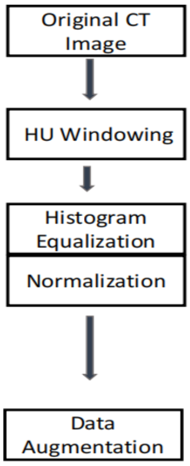

Fig. 9. Data preprocessing steps.

TABLE I. DATASETS COMPLETE INFORMATION

Dataset Parameters	3DIRCADb dataset	CHAOS dataset
CT Count	20	20
Total Slices Count	2,085	1,367
Slices count after augmentation	10,425	6,835
Acquisition phase	Portal venous	Portal venous
Image resolution	512 × 512	512 × 512
Slices per CT volume	[74–260]	[77–105]
X-Y voxel spacing (mm)	[0.56–0.87]	[0.7–0.8]
Slice thickness (mm)	[1.60–4.00]	[3–3.2]
Scanners used	Various European hospitals and different CT scanners	Philips SecuraCT with 16 detectors and a Philips Mx8000 CT with 64 detectors and Toshiba AquilionOne with 320 detectors

B. Training Parameters

The deep learning-based encoder-decoder Def-UNet needs some learning parameters to perform training. Adam optimization is used and the fixed learning rate cannot bring good convergence to the network. So we have started with the learning rate of 1×10^{-5} and when the loss stops reducing during the training process then the learning rate reduces with the decay factor of 0.1. The training continues up to 100 epochs. The beta1 and beta 2 of Adam are set to 0.9 and 0.999 after the trial and error procedure. L2 regularization with the weight penalty factor of 1×10^{-5} is fixed with the batch size of 8. The segmented pixels of medical images are generally smaller than the background pixels which causes class imbalance. The dice loss is used and this loss function can reduce the class imbalance issue present in the medical data sets. The dice loss is simply defined as a complement of the Dice coefficient, which is an evaluation metric of image segmentation, where this function calculates the similarities between the images which reduces the network weights and thus optimizes the loss of the network. The software and hardware implementations of the proposed network are: Python language is used with TensorFlow and Keras and Google Colab PRO provided the hardware GPU equipment to train and test the proposed network.

C. Evaluation Metrics

Evaluation metrics play a prominent role in evaluating the performance of the neural network. To assess the quality of the segmentation done by the proposed network we have used the following evaluation metrics. They are Dice Similarity Coefficient (DSC), Volumetric Overlap Error (VOE), Intersection of Union (IoU), Relative Absolute Volume Difference (RAVD), Average Symmetric Surface Distance (ASSD), and Maximum Symmetric Surface Distance (MSSD). The definitions and the mathematical formulas of the above-considered segmentation evaluation metrics are defined as follows:
The dice similarity coefficient calculates the overlap between the Ground Truth (GT) image and the Predicted Image (PI). The value varies between 0 and 1.

$$DSC(GT,PI) = \frac{2\times|GT\cap PI|}{|GT|+|PI|} = \frac{2\times TP}{2\times TP+FP+FN} \qquad (10)$$

Volumetric Overlap Error is the complement of the jaccard index i.e. the error calculated on it. The Jaccard index is also called the Intersection of Union (IOU).

$$VOE(GT,PI) = 1-\frac{|GT\cap PI|}{|GT\cup PI|} = 1-\frac{TP}{TP+FP+FN} \qquad (11)$$

Relative Absolute Volume Difference (RAVD): Let V_G and V_P be the volumes of ground truths and the predicted images. The difference between these two is called as relative absolute volume difference. If it returns 1 then it is called perfect segmentation and 0 means worst segmentation. It is an asymmetric metric with the formula.

$$RAVD(G,P) = \frac{|V_G-V_P|}{V_G} = \frac{FP}{TP+FN} \qquad (12)$$

ASSD and MSSD are surface metrics that calculate the correlation between the surface voxels of the predicted and ground truth images. Let SV_{GT} and SV_{PI} be the surface voxels of the ground truth and predicted image respectively and gt, pi are random voxels selected to calculate Euclidean distance d.

$$ASSD(GT,PI) = \frac{1}{|SV_{GT}|+|SV_{PI}|}$$
$$= \frac{1}{\sum_{gt\in SV_{GT}} d(gt,SV_{GT})+\sum_{pi\in SV_{PI}} d(pi,SV_{PI})} \qquad (13)$$

$$MSSD(GT,PI) = \text{MAX}\left\{\max_{gt\,\in\,SV_{GT}} d(gt,SV_{GT}),\ \max_{pi\in SV_{PI}} d(pi,SV_{PI})\right\} \qquad (14)$$

V. RESULT ANALYSIS

The learning and performance evaluation of the proposed network was done by the two datasets, CHAOS and 3DIRCADb. The combination of these two provided the complexity of the CT volume and a typical real-time scenario of the radiologist. In general, the clinician has to deal with healthy and unhealthy livers from various age groups, genders, and in particular different CT scanners. Our Def-UNet was trained in that scenario to have a more generalizable capacity and to make the network familiar with the real-time scenario. The evaluation metrics used were discussed in the previous section and the analysis table is shown in Table II. Four different experimentations were done to show the efficacy of the proposed model. (a) Network with DwSR module and pyramid pooling module. (b) network without DwSR and with pyramid pooling module (c) network with DwSR and without pyramid pooling module (d) network without DwSR and without pyramid pooling module. The main motive behind this experimentation is the know the importance of DwSR and pyramid modules and also their impact in improving the segmentation performance. Among all the segmentation metrics, the efficiency can be seen significantly from the dice coefficient metrics with values 0.974 for Global Dice Score (GDS) and 0.953 for Dice Score per Case (DSC). The liver segmentation results under the experimentation with DwSR and pyramid modules is presented in Table II. Since the proposed architecture is based on UNet, for comparision we have considered the architectures that are derived from UNet and the results are shown in Table III. No post processing techniques are used for our proposed model in order to refine the results and moreover, the proposed architecture is computationally effective with 25.6 million parameters which is very less when compared to H-DenseNet [12] with around 80 million parameters. The segmentation predictions of the model are shown in Fig. 10. The results shows the liver delineation outcomes with less segmentation error.

TABLE II. LIVER SEGMENTATION EXPERIMENTATION RESULTS

Methods	DSC	GDS	VOE	IOU	RAVD	ASSD	MSSD
With DwSR and pyramid pooling	0.953	0.974	0.058	0.938	0.067	1.988	41.684
With DwSR and without pyramid pooling	0.942	0.963	0.085	0.924	0.034	2.982	52.453
Without DwSR and with pyramid pooling	0.938	0.956	0.089	0.918	0.028	2.112	69.483
Without DwSR and pyramid pooling	0.924	0.945	0.094	0.906	0.038	3.453	81.654

TABLE III. LIVER SEGMENTATION COMPARISON RESULTS

Methods	DSC	GDS
U-Net[7]	0.892	0.913
mU-Net[18]	0.921	0.935
MA-Net[15]	0.936	0.954
CE-Net[14]	0.941	0.958
RA-Unet[19]	0.946	0.962
proposed	**0.953**	**0.974**

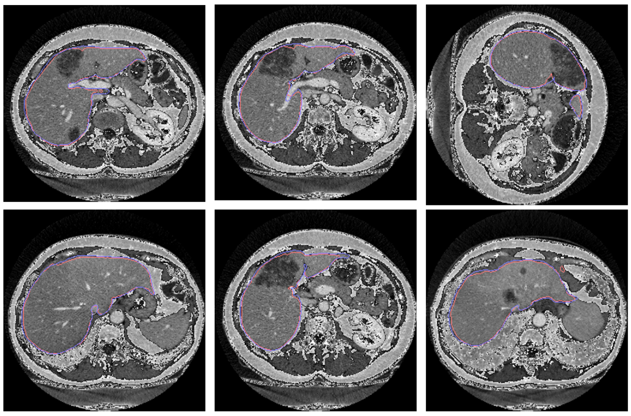

Fig. 10. Liver segmentation results. The red line shows the original segmentation and the blue line represents the predicted segmentation.

VI. CONCLUSION

In general, the radiologist has to segment healthy as well as unhealthy livers that are drawn from different types of CT scanners. Due to the design configuration of the computed tomography scanners, the CT slices vary in terms of clarity, and contrast, and even the noise artifacts also differ. Moreover, the liver contour will have many shape variants between a healthy and unhealthy patient. To resolve the liver segmentation problem from the above perspective, we have proposed a novel approach that performs liver segmentation from multi-modality CT images. On the encoder side, a deformable convolution is used to replace the normal convolution because the deformable convolution convolves based on the shape of the object with fewer parameters and also improvises the encoder information before sending it to the decoder via skip connection. The quality of the features is improved by applying feature fusion of the layer input set and the reconstructed layer output features through a recalibration and DwSR block. The pyramid pooling is used as the bottleneck layer that can extract rich contextual information even from the low-level features. We evaluated the proposed network against the other state-of-the-art methods and achieved an upliftment and comparable performance in liver segmentation with less design and computational complexities. Postprocessing techniques are not used to refine the results. Therefore, our proposed Def-UNet can be broaden to tumor segmentation or even other medical image segmentation works under various modalities.

The future work could extend the application of Def-UNet to other areas of medical image segmentation, such as tumor segmentation or segmentation tasks involving other organs and tissues across various imaging modalities. Additionally, exploring postprocessing techniques and incorporating advanced data augmentation strategies could further enhance segmentation accuracy and robustness. Investigating the integration of attention mechanisms and transfer learning could also improve the model's adaptability and performance across diverse datasets. Moreover, developing a framework for real-time segmentation and validation on larger, more diverse datasets could ensure the practical applicability of our approach in clinical settings. Lastly, collaboration with clinical experts to fine-tune the model and tailor it to specific clinical requirements would be a valuable step towards translational research and deployment in real-world healthcare environments.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

B.M.A.T. conducted the research; B.M.A.T. and S.S.B.B. analyzed the data; B.M.A.T. wrote the paper; All authors had approved the final version.

REFERENCES

[1] A. C. Gupta, G. Cazoulat, and M. A. Taie *et al.*, "Fully automated deep learning based auto-contouring of liver segments and spleen

on contrast-enhanced CT images," *Sci. Rep.*, vol. 14, no. 1, p. 4678 , 2024. https://doi.org/10.1038/s41598-024-53997-y

[2] V. Mahadevan, "Anatomy of the liver," *Surgery (Oxford)*, vol. 38, no. 8, pp. 427–431. doi: https://doi.org/10.1016/j.mpsur. 2014.10.004

[3] A. Gotra, L. Sivakumaran, and G. Chartrand *et al.*, "Liver segmentation: Indications, techniques and future directions," *Insights into Imaging*, vol. 8, pp. 377–392, 2017. doi: 10.1007/s13244-017-0558-1

[4] C. Li, X. Wang, S. Eberl, M. Fulham, Y. Yin, J. Chen, and D. D. Feng, "A likelihood and local constraint level set model for liver tumor segmentation from ct volumes," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2967–2977, 2013. doi: 10.1109/TBME.2013. 2267212

[5] X. Liu, L. Song, S. Liu, and Y. Zhang, "A review of deeplearning-based medical image segmentation methods," *Sustainability*, vol. 13, no. 3, 1224, 2021. doi: 10.3390/su13031224

[6] K. Suzuki, "Overview of deep learning in medical imaging," *Radiological Physics and Technology*, vol. 10, no. 3, pp. 257–273, 2017. doi: 10.1007/ s12194-017-0406-5

[7] P. Bilic, P. Christ, and H. B. Li *et al.*, "The liver tumor segmentation benchmark (lits)," *Medical Image Analysis*, vol. 84, 102680, 2023.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich*, Germany, October 2015, pp. 234–241.

[9] B. M, Tummala and S. S. Barpanda, "Liver tumor segmentation from computed tomography images using multiscale residual dilated encoder-decoder network," *Int. J. Imaging Syst. Technol.*, vol. 32, no. 2, pp. 600–613, 2022. https://doi.org/10.1002/ima.22640

[10] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017. doi: 10.1109/TPAMI.2016.2644615

[11] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-denseunet: Hybrid densely connected unet for liver and tumor segmentation from CT volumes," *IEEE Transactions on Medical Imaging*, vol. 37, no. 12, pp. 2663–2674, 2018. doi: 10.1109/TMI.2018.2845918

[12] C. Grzegorz, H. Meine, and J. H. Moltz *et al.*, "Neural network-based automatic liver tumor segmentation with random forest-based candidate filtering," arXiv preprint, arXiv:1706.00842, 2017.

[13] Y. Zhang, Z. He, C. Zhong, Y. Zhang, and Z. Shi, "Fully convolutional neural network with post-processing methods for automatic liver segmentation from CT," in *Proc. 2017 Chinese Automation Congress (CAC)*, 2017, pp. 3864–3869. doi: 10.1109/CAC.2017.8243454

[14] F. Lu, F. Wu, P. Hu, Z. Peng, and D. Kong, "Automatic 3D liver location and segmentation via convolutional neural network and graph cut," *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, no. 2, pp. 171–182, 2017.

[15] B. M. Tummala and S. S. Barpanda, "Curriculum learning based overcomplete U-Net for liver tumor segmentation from computed tomography images," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 3, pp. 1620–1629, 2023. https://doi.org/10.11591/eei.v12i3.4676

[16] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1856–1867, 2019.

[17] H. Seo, C. Huang, M. Bassenne, R. Xiao, and L. Xing, "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*, vol. 39, no. 5, pp. 1316–1325, 2020. doi: 10.1109/TMI.2019.2948320

[18] T. Fan, G. Wang, Y. Li, and H. Wang, "Ma-Net: A multi-scale attention network for liver and tumor segmentation," *IEEE Access*, vol. 8, pp. 179656–179665, 2020. doi: 10.1109/ACCESS.2020.3025372

[19] X. Wang, S. Han, Y. Chen, D. Gao, and N. Vasconcelos, "Volumetric attention for 3D medical image segmentation and detection," in *Proc. Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference*, Shenzhen, China, October, 2019, pp. 175–184. doi: 10.1007/978-3-030-32226-7 20

[20] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 2016 FOUrth International Conference on 3D Vision (3DV)*, 2016, pp. 565–571.

[21] Z. Gu, J. Cheng, and H. Fu *et al.*, "Ce-net: Context encoder network for 2d medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 38, pp. 2281–2292, 2019. doi: 10.1109/TMI.2019.2903562

[22] Q. Jin, Z. Meng, C. Sun, H. Cui, and R. Su, "Ra-unet: A hybrid deep attention-aware network to extract liver and tumor in CT scans," *Frontiers in Bioengineering and Biotechnology*, vol. 8, no. 12, 2020. doi: 10.3389/fbioe.2020.605132

[23] L. Teng, H. Li, and S. Karim, "Dmcnn: A deep multiscale convolutional neural network model for medical image segmentation," *Journal of Healthcare Engineering*, vol. 2019, no. 12, pp. 1–10. doi: 10.1155/2019/8597606

[24] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141. doi: 10.1109/CVPR.2018.00745

[25] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proc. the European Conference on Computer Vision (ECCV)*, 2018, pp. 3–19.

[26] L. Soler *et al.* (2010). 3d image reconstruction for comparison of algorithm database: A patient specific anatomical and medical image database. [Onilne]. Available: https://www.ircad.fr/research/ data-sets/liver-segmentation-3d-ircadb-01/

[27] A. E. Kavur, M. A. Selver, O. Dicle, M. Bar, and N. S. Gezer, "CHAOS-combined (CT-MR) healthy abdominal organ segmentation challenge data," *Medical Image Analysis*, vol. 69, 101950, 2019. doi: 10.5281/ZENODO.3431873

[28] S. Pizer, R. Johnston, J. Ericksen, B. Yankaskas, and K. Muller, "Contrastlimited adaptive histogram equalization: Speed and effectiveness," in *Proc. the First Conference on Visualization in Biomedical Computing*, Los Alamitos, CA, USA, 1990, pp. 337–345. doi: 10.1109/VBC.1990.109340