IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Journal Européen des Systèmes Automatisés
Vol. 57, No. 6, December, 2024, pp. 1775-1783
Journal homepage: http://iieta.org/journals/jesa

Enhancing Liver Tumour Segmentation in CT Images Using Dilated Residual Capsule Networks

Jyoshna Allenki*, Hemant Kumar Soni

Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University Madhya Pradesh, Gwalior 474005, India

Corresponding Author Email: allenkijyoshna@gmail.com

https://doi.org/10.18280/jesa.570625

Received: 24 July 2024
Revised: 19 November 2024
Accepted: 29 November 2024
Available online: 31 December 2024

Keywords:
liver CT images, dilated convolutional residual networks, capsule network, ResNet-50, Multiple Dilated Convolution Block (MDCB), parameterized ReLU (PReLU), USAT-Nets++, ED-SwinNets++

ABSTRACT

Liver CT images play a crucial role in the early diagnosis of liver disorders and have proven effective in identifying chronic liver disease, which may lead to fatal outcomes. This imaging technique provides detailed cross-sectional views, allowing for precise detection of abnormalities, aiding in timely intervention, and improving patient prognosis. The detection process of chronic liver disease should be carried out with meticulous accuracy. Due to the inherent complexities involved and the presence of ambiguities in CT images, segmentation approaches have not yet reached the pinnacle of accurate and reliable performance required for clinical application. Recently, the emergence of machine learning and deep learning algorithms has provided valuable insights into achieving a more accurate segmentation process. However, these existing deep learning algorithms suffer from several challenges that hinder segmentation performance. Hence, independent deep learning algorithms require further refinement to handle CT liver images effectively. To address this problem, this research article proposes a fully automated, robust, and accurate segmentation of CT liver images based on a deep neural network architecture that adopts dilated residual networks integrated with powerful capsule networks. This proposed network combines the strengths of capsule networks and ResNet-50 architectures to achieve better segmentation results. Extensive experimentation is conducted using 100 healthy subjects, and 131 contrast-enhanced image data are used for training, while 70 CT images are used for testing. Furthermore, the proposed model is evaluated using performance metrics such as DICE, Intersection over Union (IoU), precision, and recall. To demonstrate the superiority of the suggested network, its segmentation performance is compared with that of existing state-of-the-art deep learning architectures. The results demonstrate that the suggested model achieved 0.98 DICE, 0.95 IoU, 99.2% precision, and 99.1% recall, respectively, surpassing various existing models used for liver CT image segmentation.

1. INTRODUCTION

The liver, the largest gland in the human body, weighs roughly 1500 grams. It plays a critical role in metabolism, digestion, and detoxification. Hepatic steatosis and hepatitis are significant liver disorders that can lead to hepatic sclerosis and liver cancer. Liver tumors are a leading cause of cancer-related deaths, accounting for approximately 841,080 cases and 781,631 deaths globally in 2020 [1, 2]. Hence, an intelligent system is required for accurately locating tumor areas in the liver [3].

For the assessment and staging of liver tumors, computed tomography (CT) is one of the primary popular imaging modalities [4, 5]. Typically, skilled radiologists manually delineate the liver and liver tumor segments. However, manually tracing volumetric CT images slice-by-slice is labor-intensive, subjective, and not very efficient.

Automated or semi-automated segmentation methods would significantly increase efficiency. Moreover, the need for automated liver and liver tumor segmentation is emphasized by the growing use of intraoperative 3D imaging systems [6].

Moreover, machine learning methods like Artificial Neural Networks (ANN) and Support Vector Machines (SVM) are frequently employed for liver tumor segmentation and feature extraction [7-10]. Although both are considered black-box models, they offer valuable characteristics such as parallel processing and data transformation through kernel functions. These techniques yield highly accurate classification results, particularly in the segmentation process. However, prior to the classification steps such as pre-processing, semantic segmentation and feature extraction were followed to achieve the high detection rate. Recently, researchers shown huge interest in adopting the deep learning networks for semantic segmentation of Liver tumours. The shift from traditional machine learning frameworks to deep learning (DL) is driven by the remarkable accuracy attained through its extensive learning architectures. These structures enable DL to extract more intricate features from the data. Existing DL methods such as U-NETS [11], SegNets [12], Fully Connected

Convolutional Neural Networks [13] and Hybrid ResNets [14, 15] are deployed for the segmentation. Furthermore, these methods suffer from the gradient problems [16-20] which stops to achieve the best performance. Inspired by this issue, this research paper presents the efficient ensemble of capsule and dilated residual networks for achieving the best segmentation. To sum up the advantages, the primary contribution of this paper is summarized as pursues:

1. Proposing an intelligent framework for the effective segmentation of Liver tumours based on CT liver images;
2. Proposing the novel capsule based dilated residual networks which propels the semantic segmentation to achieve its peak performance that can aid for the better classification;
3. Investigating the proposed model by measuring the various performance metrics and comparing with the alternate state-of-art existing learning models.

The remainder of the content is organized as pursues: Section 2 presents the different segmentation techniques demonstrated by different authors. The proposed methodology and dataset descriptions are depicted in Section 3. The experimental outcomes are demonstrated in Section 4. The paper is ultimately wrapped up with prospects for future improvements delineated in Section 5.

2. RELATED WORK

For an efficient liver segmentation process, shape prior methods and anatomical knowledge about the organs are included. Many works have demonstrated the effectiveness of designing model-based segmentation methods. These approaches leverage a detailed understanding of liver anatomy to improve accuracy and reliability in segmentation.

Rahman et al. [21] introduced a liver segmentation pipeline utilizing graph-cut techniques in a 3D interactive environment. Wang et al. [22] utilized a combination approach, amalgamating an expectation maximization algorithm with region-based texture classification for segmentation.

Despite their limitations, such methods might struggle with distinguishing the liver from surrounding tissues due to poor contrast and indistinct boundaries. Additionally, the use of contrast agents to enhance tumor visibility in CT scans introduces noise. Recently, deep learning techniques, particularly fully convolutional networks (FCNs), have demonstrated significant promise in automatically segmenting medical images.

Shao et al. [23] introduced a novel deep learning approach, termed a 3D convolutional neural network (CNN), for the automated segmentation of livers. This method trains to generate subject-specific probability maps, serving as shape priors to establish the initial liver surface. The model integrates both local and global information to refine segmentation. Global data encompass healthy liver regions, capturing area appearance and intensity distribution, while local nonparametric data focus on detecting abnormal liver features. Zhang et al. [24] presented a dual-stage mechanism, in which the liver was segmented first by utilizing shape models, followed by tumor segmentation using dense random trees with auto-context learning schemes. The proposed scheme suffers from low performance and significantly consumes more computational overhead.

Wu et al. [25] proposed ensemble dense networks by integrating the 2D-U-Net and 3D-convolutional neural networks to achieve better performance. However, the persistence of vanishing gradient problems still prevents these models from achieving better segmentation accuracies. To overcome these issues, Manjunath et al. [26] suggested the enhanced U-NET by integrating the residual path into the skip connection of the U-NET. This method provides good performance but still requires improvements in terms of computational overhead. Additionally, Jiang et al. [27] presented Convolutional Neural Networks that incorporate attention maps and skip connections.

3. PROPOSED WORK

The entire structure of the suggested model is depicted in Figure 1, comprising three constituent elements: capsule network, dilated residual networks, and residual skip connection. The components are constructed with a U-shaped architecture, which is very similar to U-Nets. The three components are integrated as the encoder-decoder framework. The capsule networks and down-samplers are used in the encoder, whereas the up-sampling with dilated residual networks is used as the decoder. As the first step, features are extracted and fed to the encoders, which consist of the capsule network coupled with down-sampling. Following the encoder design, the features are then fed to the decoders coupled with the dilated residual block with up-sampling. The skip connection serves to link the encoder and decoder blocks for interconnection purposes. Finally, all the features are concatenated to form the segmented images. The depiction of the suggested network is thoroughly elucidated in the prior segment.

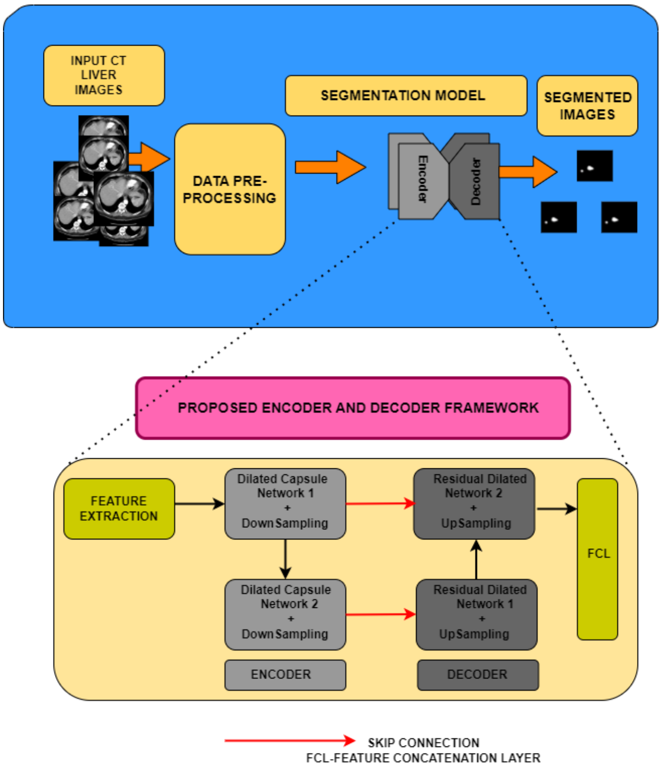
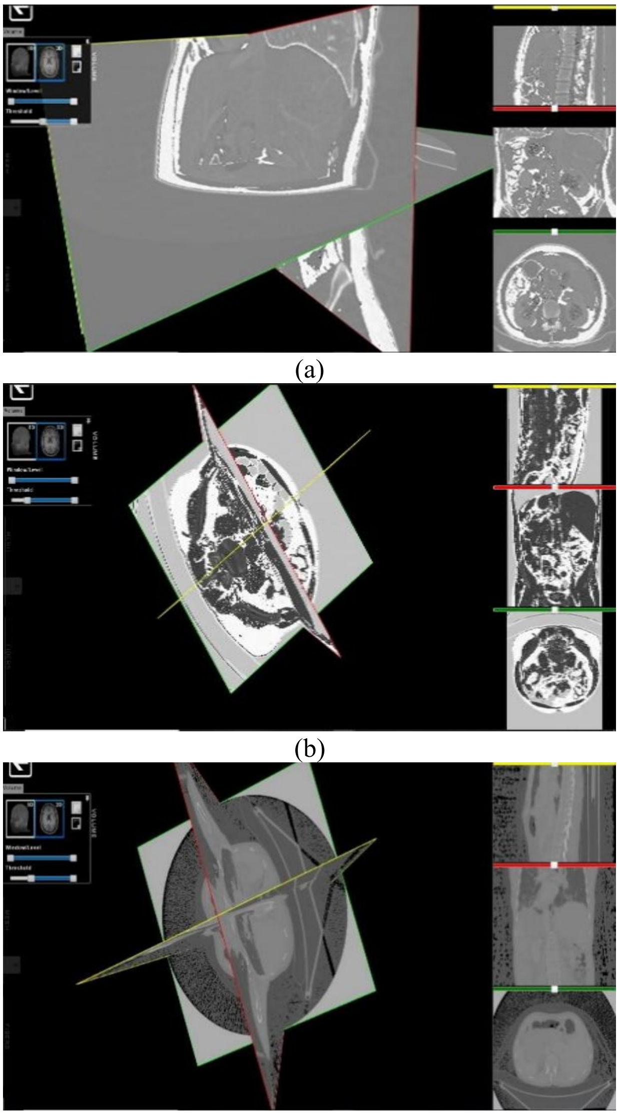

INPUT CT LIVER IMAGES
DATA PRE-PROCESSING
SEGMENTATION MODEL
Encoder
Decoder
SEGMENTED IMAGES
PROPOSED ENCODER AND DECODER FRAMEWORK
FEATURE EXTRACTION
Dilated Capsule Network 1 + DownSampling
Dilated Capsule Network 2 + DownSampling
Residual Dilated Network 2 + UpSampling
Residual Dilated Network 1 + UpSampling
FCL
ENCODER
DECODER
SKIP CONNECTION
FCL-FEATURE CONCATENATION LAYER

Figure 1. Proposed architecture for segmenting the liver tumors using LiTS liver images

3.1 Methods and materials

The dataset used for evaluating the proposed framework is the LiTS-2017 liver tumor segmentation model. Nearly 131 contrast-enhanced images are utilized for training, while 70 CT images are used for testing, with all images having a resolution of 300 dpsi. These datasets exhibit heterogeneous and diffuse shapes and are organized in conjunction with ISBI2017 and MICCAI 2017 datasets. Figure 2 presents the sample images used for the evaluation process. As shown in Figure 2, all the images are stored in (.nii format), and Python code is deployed to read these files.



(a)

(b)

(c)

Figure 2. Sample chronic liver images (.nii format) from the LiTS datasets used for segmentation and classification model

3.2 Data pre - processing technique

The medical preprocessing method aims to eliminate noisy and low-quality pixels that can impede the accurate detection of liver cancers in CT scan images. Pixel-intensive evaluation has been applied to eradicate inconsistent and noisy pixels from the input images. Additionally, image histogram techniques have been employed to improve image quality, as they demonstrate superior efficacy across various image types.

3.3 Core model design

The core model design consists of capsule networks, dilated networks, and finally encoder-decoder design. The detailed description of each component as follows:

3.3.1 Capsule networks-its working background

In the realm of deep learning, Convolutional Neural Networks (CNNs) have garnered significant attention for tackling image visual representation. Nonetheless, these CNNs encounter notable limitations in their fundamental architecture, resulting in subpar performance across various tasks. CNNs autonomously discern features from images, which is crucial for detecting and recognizing diverse visual objects. Initially, they identify simple features like edges. As the layers go deeper, they recognize more complex features. However, CNNs primarily focus on extracting features without adequately considering the spatial arrangement of objects, which is a significant architectural flaw. Given the necessity for precise feature extraction in thermal image inputs, CNNs require modifications to enhance accuracy. Addressing this concern, the pooling layers in CNNs are substituted with capsule networks to improve spatial feature extraction. This research introduces the concept of fast capsule networks aimed at enhancing spatial feature extraction for improved performance.

Five convolutional layers are employed for feature extraction, which is then integrated with the capsule network to capture the spatial details of DFU. The complete network utilizes the ReLU activation function throughout.

Capsule Networks were recently introduced to tackle the shortcomings of conventional CNN architectures. Capsules represent clusters of neurons encoding both spatial details and the likelihood of object presence. Within a capsule network, every element in an image is associated with a capsule, providing encoded information about its presence and spatial characteristics.

1. The likelihood of presence within entities.
2. Parameters for the instantiation of entities.

The capsule network comprises 3 CNN layers. This network is segmented into three tiers: a base capsule layer, an upper capsule layer, and a classification tier. Global parameter sharing is executed to minimize error accumulation, while employing an optimized dynamic routing algorithm for parameter updates iteratively. To encode the crucial spatial correlation between low and high-level convolutional characteristics in the image, the product of the input vector matrix with the weight matrix is computed.

$$Y(i.j) = W_{i,j}U(i,j) * S_j \tag{1}$$

where, *Y*(*i*,*j*) represents the product output of the input vector matrix with the weight matrix for the capsule network, W(*i*,*j*) is the weight matrix between capsule *i* and capsule *j*, *U*(*i*,*j*) is the input vector matrix for capsule *i* and capsule *j* and *Sj* is the output of capsule *j*, which is computed through a combination of weighted input vectors.

To ascertain the current capsules, their weighted input vectors are summed up, directing the resulting output to the superior level capsule.

$$S(j) = \sum_{j} Y(i,j) * D(j) \tag{2}$$

where, *S*(*j*) represents the summed weighted input vector for capsule *j*, *Y*(*i*,*j*) is the product of the input vector matrix with the weight matrix and *D*(*j*) represents the dynamic routing coefficient for capsule *j*, which adjusts the weight based on the routing process.

The application of non-linearity is culminated by employing the squash function

$$G(j) = squash(S(j)) \tag{3}$$

where, *G*(*j*) represents the output of capsule *j* after applying the squash function.

For precise segmentation, capsule networks have the capability to capture data situated across various locations and discern the correlations between features through the utilization of mathematical Eq. (1).

The convolutional layers reside within the lower capsule area, while the primary capsules occupy the upper region. Eq. (2) is utilized to compute the output weights, which are then transmitted to the upper capsule region. The squash function maintains the original vector direction by compressing its length within the range of (0, 1). Subsequently, the model integrates dot product operations among similar capsules and optimizes dynamic routing to generate output. This iterative process continually adjusts network weights to construct feature maps. Following this, the dimensionality of the feature maps is reduced through fully connected layers and transformed into one-dimensional feature maps via flattening layers. The one-dimensional capsule matrix *G* undergoes flattening to yield a one-dimensional vector *P*, which is further transformed into a vector of length L through a fully connected layer. Eq. (4) represents the mathematical formulation for these features.

$$Z = F(G(j), P) \tag{4}$$

3.3.2 Dilated residual networks

The dilated residual network is constructed based on the ResNet-50 model. It consists of 18 layers and all the convolutional layers are replaced with dilated convolution layers (DCL). The Dilated Convolution layer involves introducing gaps (zero padding) between the components of the convolutional layers to enlarge the convolutional kernel, as referenced in the literature. Let's denote the expansion factor of the dilated convolution as variable t, which is expressed mathematically as

$$C = C + (d - 1)(t - 1) \tag{5}$$

The C represents the dilated convolution blocks subsequent to integrating the perforations, while c denotes the initial convolutional kernel. The expansion rate, t, also serves as a measure of the original convolutional kernel's expansion capacity. Receptive fields are employed to delineate the distinct dilation processes. Consequently, dilated convolution layers (DCN) have garnered significant attention in research, proving highly effective in extracting spatial features without sacrificing computational efficiency or imposing additional computational overhead. However, the utilization of convolution blocks with identical dilation rates may still give rise to the gridding phenomenon. To overcome the gridding problems, it is found that the combination of different dilation rates can reduce the gridding phenomenon with the rich extraction of features from the medical images. Hence, Multiple Dilated Convolution Block (MDCB) module is constructed using MDCB which can aid for rich extraction of features. As depicted in Figure 3, it comprises Convolutional layer-1, succeeded by the residual units and attention maps. The initial Convolutional layer employs 8 kernel filters sized 3×3, subsequently followed by batch normalization (BN) and an activation layer (LReLU). In contrast, residual units 2 and 3 incorporate a convolutional block succeeded by an identity block. The proposed study adopts residual units to address the issue of gradient vanishing. The design of the identity block mirrors that of the convolutional block. Finally, a Max-Average pooling (Average Pooling) layer transforms 2D feature maps into 1D feature maps.

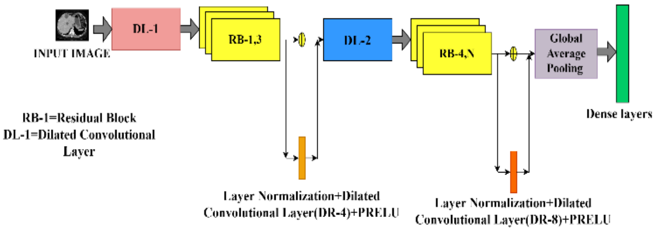

INPUT IMAGE
DL-1
RB-1,3
DL-2
RB-4,N
Global Average Pooling
Dense layers
RB-1=Residual Block
DL-1=Dilated Convolutional Layer
Layer Normalization+Dilated Convolutional Layer(DR-4)+PRELU
Layer Normalization+Dilated Convolutional Layer(DR-8)+PRELU

Figure 3. Multi dilated residual convolutional network for an effective segmentation

3.4 Segmentation model

The core segmentation model which depends on the U-Net architecture depicted in Figure 4. This architecture can be seen as the extension of U-Net and combined it with the designed module and networks. The ED-SwinNets++ consist of four parts: (a) feature extraction, (b) modified Capsule, (c) embedded encoder (Capsule-Encoder) and (d) Decoder (DCL-Decoder), skip connections up-sampling and down-sampling blocks. A significant benefit of the suggested framework lies in its capacity to extract intricate underlying characteristics and enhance the U-nets' expressive efficacy. Integrating the convolutional attention module can additionally enhance the encoders' proficiency in capturing pertinent contextual features while suppressing irrelevant ones. Conversely, the utilization of DCL within decoders will merge low-level features with high-level ones, thereby augmenting segmentation performance without any overlap.

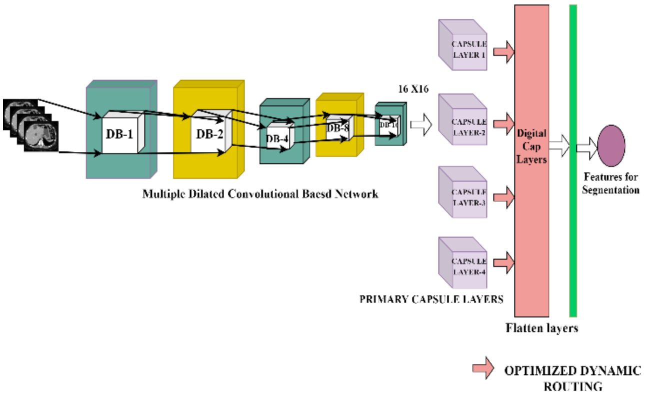


Figure 4. Multi dilated convolutional layers based capsule networks

3.5 Feature extraction process

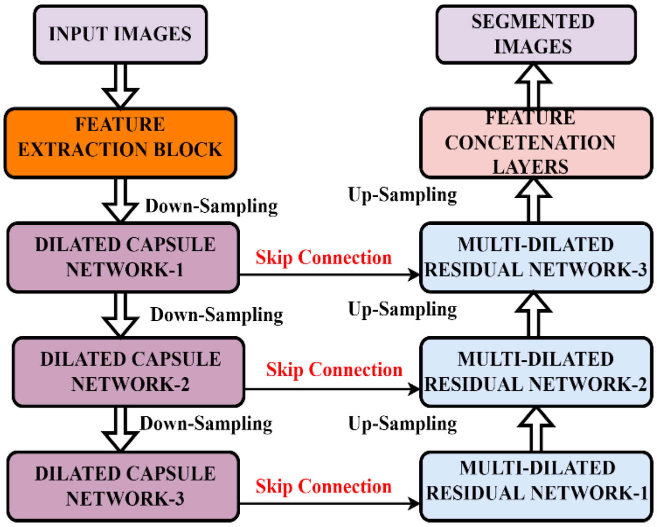
The primary role of the feature extraction module revolves around transforming individual input images (I) into high-dimensional tensors (T) with dimensions represented as

$$TI \in RL/4 \times B/4 \times \frac{C}{4} \quad (6)$$

where L, B, and C denote the height, width, and sequence length of each input patch, respectively. In contrast to other works, the proposed layer is constructed with four dilated consecutive convolutional layers and uses PReLU activation functions in place of ReLU. Also, layer normalization is adopted for each layer. The addition of the PReLU activation function enables the proposed network to achieve high accuracy in the spatial extraction of features with less computational overhead.

3.6 Encoder design

Figure 5 shows the encoder–decoder design based on the USAT-Nets++. The extracted characteristics are inputted into the suggested encoder, comprising four phases, with each phase housing a proposed capsule network alongside down-sampling. As the network progresses deeper, the quantity of characteristics will be diminished to generate a hierarchical representation. Within the initial three phases, the inputted characteristics will undergo a concatenation process to lower feature resolution and enhance dimensionality following the transformations of the proposed networks and the down-sampling network.



INPUT IMAGES
FEATURE EXTRACTION BLOCK
Down-Sampling
DILATED CAPSULE NETWORK-1
Skip Connection
Down-Sampling
DILATED CAPSULE NETWORK-2
Skip Connection
Down-Sampling
DILATED CAPSULE NETWORK-3
Skip Connection
MULTI-DILATED RESIDUAL NETWORK-1
Up-Sampling
MULTI-DILATED RESIDUAL NETWORK-2
Up-Sampling
MULTI-DILATED RESIDUAL NETWORK-3
Up-Sampling
FEATURE CONCETENATION LAYERS
SEGMENTED IMAGES

Figure 5. Encoder-decoder architecture for image segmentation

3.7 Decoder design

The decoder primarily comprises three phases. Unlike prior iterations of the U-net model, the suggested decoder integrates the novel transformer block before executing up-sampling and incorporating skip connections. More precisely, the encoder's output serves as the decoder's input. Within each decoder stage, the input characteristics undergo a 2x up-sampling, subsequently merging with skip connection feature maps from the corresponding encoder stage before being directed to the proposed MDCB layers.

After completing the aforementioned three stages, we obtain the output with a resolution of L/4 × H/4. Utilizing a 4× up-sampling operator directly would result in the loss of numerous shallow features. Hence, we opt for down-sampling the input image by amalgamating two blocks to acquire low-level features with resolutions of L×B and L/2×B/2. Each block comprises a 3×3 convolutional layer, a group normalization layer, and a PReLU layer consecutively. These output features are then utilized to derive the final mask predictions via skip connections. Following a methodology akin to U-Net, skip connections are leveraged to amalgamate multiscale features from the encoder with up-sampled features from the decoder.

3.8 Feature concatenation layer

After gathering the attributes from the dual-branch encoder-decoder, a feature concatenation layer (FCL) is utilized to concatenate the multi-scale features. To perform the concatenation operation in an accurate manner, a transformer block is constructed by utilizing the MHCSAM to facilitate productive communication among the various hierarchical characteristics to form the segmented image. In this process, tokens are generated from the features formed, and then convolutional self-attention maps are computed for each token, reshaped by another layer. In these computations, only two layers of CSAM are needed, which means computational complexity is significantly reduced compared to the conventional straightforward self-attention maps. Then, the output concatenated maps are calculated as follows:

$$G_n'' = Transfomer(Flatten(GAPL(G'(n)))) \quad (7)$$

$$G^{Output} = [G_1'', G_2'', G_n''] \quad (8)$$

where, G_1'' is the tokens created from the feature maps, which represents the total global abstract information from every features.

For extracting the global features, Flatten and Global Average Pooling layers (GAPL) are used and fed to transformers for calculating the concatenated features. Hence the proposed model facilitates proficient amalgamation of features across multiple scales, consequently leading to enhanced segmentation efficacy.

4. IMPLEMENTATION DETAILS

The network framework suggested in this investigation was constructed utilizing Keras, with TensorFlow serving as the backend. Table 1 delineates the hyperparameters employed in the model's training process.

Table 1. Training parameters for the proposed algorithm [1]

Sl. No.	Hyperparameters Used	Specifications
1	Initial learning rate	0.001
2	No of Epochs used	150
3	Batch size	30
4	Optimizer	ADAM
5	Momentum	0.12

During the training phase, an early termination technique was employed to halt the training prematurely, mitigating the risk of overfitting. Diverse augmented images were utilized to enrich the training regimen. The entire algorithm underwent experimentation on a personal computer workstation equipped with 16 gigabytes of RAM, a 2-terabyte solid-state drive, an Intel i7 processor, an NVIDIA GeForce RTX graphics card, and a 3.4 gigahertz operating frequency. And used Python, TensorFlow software.

4.1 Evaluation metrics

To evaluate the efficiency of the implemented algorithm, DICE, IoU, non-biased accuracy, precision, and F1-score are utilized. Table 2 presents the mathematical formulae for computing these metrics.

Table 2. Mathematical expressions for calculating the segmentation metrics [2]

Sl. No	Performance Metrics	Expression
1	DICE (DSC)	$2(\lvert S\cap T\rvert)/(\lvert S\rvert+\lvert T\rvert)$
2	IOU	$(S\cap T)/S\cup T$
3	Peak-to-Signal Noise Ratio (PSNR)	$\lvert S\cap T\rvert/S$
4	SSIM	$S\cap T/\lvert S\rvert$

In this context, S signifies the authentic data, while T denotes the output from the model's predictions. When considering DSC and IoU, the spectrum extends from 0 to 1, where 0 indicates no intersection and 1 signifies flawless segmentation. Higher values in these metrics indicate a greater degree of overlap between the model's predictions and the ground truth, reflecting increased similarity and improved segmentation.

The paper employs the early stopping technique to tackle the problem of overfitting in the network and uplift its generalization capability. This approach enables the termination of the network training process when there is no observable enhancement in the validation performance for N consecutive instances, thereby mitigating overfitting and enhancing generalization.

4.2 Result analysis

In this section, the proposed model was compared with the alternate advanced deep learning models in phase of performance. In this evaluation, advanced deep learning structures such as Resnets [28], U-Nets [29], EfficiNets + DenseNets [30], ShuffleNets [31], EfficeiNets + ShuffleNets [32], and DE-ResNets [33] are utilized for comparison. It's important to note that every model underwent training under identical experimental conditions, utilizing datasets [34] prepared and metrics specified in the Table 3 for examination and comparison. The trained models underwent validation and assessment using test data. Across all scenarios, datasets were partitioned into 80% for training and 20% for testing.

Figure 6 shows the segmentation results for sample thermal images, comparing the ground truth with the predicted outputs. It illustrates the model's accuracy in identifying key features in the images.

Table 3. Mathematical expressions for the performance metrics' calculation [17]

Sl. No.	Performance Metrics	Mathematical Expression
1	Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
2	Recall	$\frac{TP}{TP+FN}\times 100$
3	Specificity	$\frac{TN}{TN+FP}$
4	Precision	$\frac{TN}{TP+FP}$
5	F1-Score	$2\times\frac{Precison * Recall}{Precision+Recall}$

Note: TP is True Positive Values, TN is True Negative Values, FP is False Positive and FN is False negative values

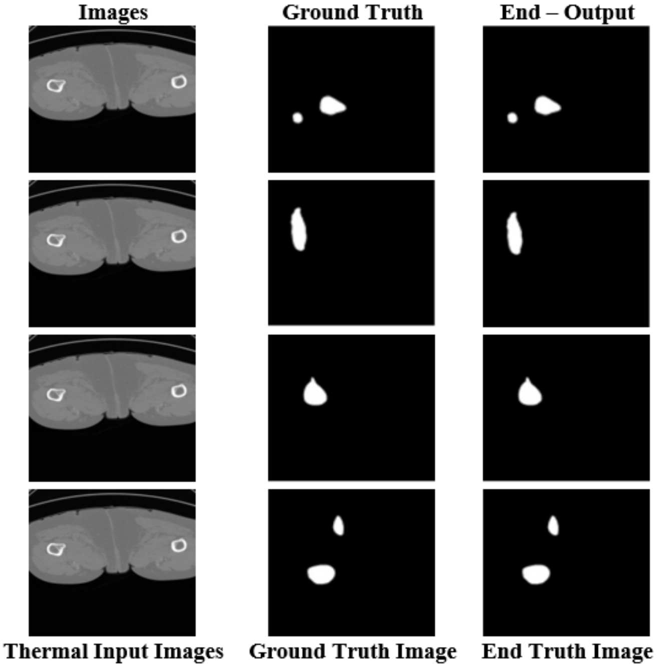


Figure 6. Segmentation outcomes for the sample thermal images with the ground truth and predicted outputs

4.3 Discussion

Tables 4-7 display a comparative examination of the average results generated by various algorithms utilizing LiTS datasets generated in Section 3.1.

Table 4. Average performance metrics of the state-of-the-art methods for thermal image segmentation process at testing phase

Algorithms	Performance Metrics			
	PSNR (dB)	SSIM	DICE	IoU
ResNets	32.5	0.74	0.73	0.6
U-Nets	30.45	0.72	0.82	0.57
EfficientNets+DenseNets	30.4	0.70	0.84	0.62
ShuffleNets	30.4	0.75	0.87	0.65
EfficientNets+ShuffleNets	32.5	0.79	0.86	0.75
DE-ResNets	33.56	0.84	0.89	0.79
Proposed Model (Dilated Residual Capsule Network)	**35.02**	**0.91**	**0.92**	**0.95**

Table 5. Average performance metrics of the state-of-the-art methods for thermal image segmentation process at validation phase

Algorithms	Performance Metrics			
	PSNR (dB)	SSIM	DICE	IoU
ResNets	32.5	0.74	0.73	0.6
U-Nets	30.45	0.72	0.82	0.57
EfficientNets + DenseNets	30.4	0.70	0.84	0.62
ShuffleNets	30.4	0.75	0.87	0.65
EfficientNets + ShuffleNets	32.5	0.79	0.86	0.75
DE-ResNets	33.56	0.84	0.89	0.79
Proposed Model (Dilated Residual Capsule Network)	**35.02**	**0.91**	**0.92**	**0.95**

Table 6. Average performance metrics of the cutting-edge techniques for the thermal image segmentation at testing phase

Algorithms	Performance Metrics			
	Accuracy	Precision	Recall	F1-Measure
ResNets	0.84	0.84	0.83	0.67
U-Nets	0.80	0.78	0.82	0.54
EfficientNets + DenseNets	0.89	0.81	0.85	0.5
ShuffleNets	0.87	0.82	0.85	0.56
EfficientNets + ShuffleNets	0.86	0.84	0.87	0.52
DE-ResNets	0.95	0.87	0.88	0.73
Proposed Model (Dilated Residual Capsule Network)	**31.8**	**0.93**	**0.92**	**0.9**

Table 7. Average performance metrics of the cutting-edge techniques for the thermal image segmentation at validation phase

Algorithms	Performance Metrics			
	Accuracy	Precision	Recall	F1-Measure
ResNets	0.84	0.84	0.83	0.67
U-Nets	0.80	0.78	0.82	0.54
EfficientNets + DenseNets	0.89	0.81	0.85	0.5
ShuffleNets	0.87	0.82	0.85	0.56
EfficientNets + ShuffleNets	0.86	0.84	0.87	0.52
DE-ResNets	0.95	0.87	0.88	0.73
Proposed Model (Dilated Residual Capsule Network)	**31.8**	**0.93**	**0.92**	**0.9**

The suggested methodology's visual representations make it clear that it has surpassed other preexisting models by achieving superior segmentation performance. From the analysis, proposed model and DE-ResNets has produced the similar performances and other models also produced the considerable better performance in testing scenario. The DICE metric of the suggested methodology demonstrates noticeable findings of 0.92 and average segmentation performance is found to be 0.94 which is far better than the other models such as Resnets, U-Nets, EfficiNets + DenseNets, ShuffleNets, EfficeiNets + ShuffleNets and DE-ResNets. Obviously, integration of MHCSAM layer in suggested Swin Transformer architecture has further enhanced the generation of sharper and well-defined images of liver tumours by injecting additional resources. From the Table 6 shows the different computational evaluation for the various deep learning methodologies used for segmenting the liver tumour images using thermal images. It is observed from the Table 6, it is evident that the suggested methodology consumes only 12.0 secs in segmenting the thermal images and it is due to that the suggested technique contains of MHCSAM that demonstrated its indispensable function in attaining optimal segmentation efficacy 20% higher than the existing model.

4.4 Ablation experiments

In this part, ablation experimentation is carried out to demonstrate the efficacy of each element within the proposed Swin framework, we undertake ablation investigations to assess the impact of diverse factors on our model. Additionally, we conduct experiments on four different iterations of the proposed architecture to gauge their respective influences. Table 8 presents a comparative analysis of the computational costs associated with various deep learning algorithms for segmenting thermal images

Table 8. Computational cost comparison for the deep learning algorithms in segmenting thermal images

Deep Learning Algorithms	Computational Cost (MB)
ResNets	23
U-Nets	20
EfficientNets + DenseNets	23
ShuffleNets	19
EfficientNets + ShuffleNets	18.5
DE-ResNets	14.5
Proposed Model (Dilated Residual Capsule Network)	**12.0**

Table 9 shows the ablation analysis of the different variants of deep learning techniques in which the suggested methodology yields the superior performance beyond all the variants. The suggested framework has demonstrated superior mDICE and mIoU metrics compared to alternative architectures. In general, integration of CSAM in Swin Transformer block and inclusion of the feature concatenation layer has shown the promising results in achieving the segmentation performance.

Table 9. The outcomes of segmentation across various algorithms subsequent to the ablation investigation utilizing the test dataset

Algorithms	PSNR	SSIM	mDSC (%)	mIoU (%)
ResNets	32.5	0.74	0.73	0.6
U-Nets	30.45	0.72	0.82	0.57
EfficientNets+DenseNets	30.4	0.70	0.84	0.62
ShuffleNets	30.4	0.75	0.87	0.65
EfficientNets+ShuffleNets	32.5	0.79	0.86	0.75
DE-ResNets	33.56	0.84	0.89	0.79
Proposed Model (Dilated Residual Capsule Network)	**35.02**	**0.91**	**0.92**	**0.95**

5. CONCLUSIONS

In this work, an encoder-decoder-based U-shaped Swin Transformer framework is proposed for thermal liver CT image segmentation. The proposed model is based on the hierarchical representation of features. The dual-branch Swin transformer (encoder-decoder) is also innovatively added to extract the multi-scale features, in which the multi-headed convolutional self-attention layers are replaced by traditional self-attention layers. This reduces the complexity and also extracts the multi-scale features from the images. Moreover, a feature concatenation layer is embedded as the fusion module in the proposed architecture to build the long-range dependencies, allowing features to be effectively concatenated for attaining a superior segmentation mechanism. Extensive experiments on the thermogram image datasets are conducted, and performance metrics such as DICE, PSNR, SSIM, and IoU are measured and evaluated. Results demonstrate that the suggested methodology significantly surpassed alternative deep learning architectures. As the future scope, the proposed model needs its improvisation by designing more lightweight transformers and better learning the structural pixel-level features for segmenting the thermal images.

REFERENCES

[1] Rodimova, S.A., Kuznetsova, D.S., Bobrov, N.V., Gulin, A.A., Vasin, A.A., Gubina, M.V., Scheslavsky, V.I., Elagin, V.V., Karabut, M.M., Zagainov, V.E., Zagaynova, E.V. (2021). Multiphoton microscopy and mass spectrometry for revealing metabolic heterogeneity of hepatocytes in vivo. Modern Technologies in Medicine, 13(2): 18-39. https://doi.org/10.17691/stm2021.13.2.02

[2] Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 71(3): 209-249. https://doi.org/10.3322/caac.21660

[3] Hendriks, P., Boel, F., Oosterveer, T.T., Broersen, A., de Geus-Oei, L.F., Dijkstra, J., Burgmans, M.C. (2023). Ablation margin quantification after thermal ablation of malignant liver tumors: How to optimize the procedure? A systematic review of the available evidence. European Journal of Radiology Open, 11: 100501. https://doi.org/10.1016/j.ejro.2023.100501

[4] Taunk, N.K., Burgdorf, B., Dong, L., Ben-Josef, E. (2021). Simultaneous multiple liver metastasis treated with pencil beam proton stereotactic body radiotherapy (SBRT). International Journal of Particle Therapy, 8(2): 89-94. https://doi.org/10.14338/IJPT-20-00085.1

[5] Reddy, V.N., Rao, P.S. (2018). Comparative analysis of breast cancer detection using K-means and FCM and EM segmentation techniques. Ingenierie des Systemes d'Information, 23(6): 173-187.

[6] Midya, A., Chakraborty, J., Srouji, R., Narayan, R.R., et al. (2023). Computerized diagnosis of liver tumors from CT scans using a deep neural network approach. IEEE Journal of Biomedical and Health Informatics, 27(5): 2456-2464. https://doi.org/10.1109/JBHI.2023.3248489

[7] Hirra, I., Ahmad, M., Hussain, A., Ashraf, M.U., Saeed, I.A., Qadri, S.F., Alghamdi, A.M., Alfakeeh, A.S. (2021). Breast cancer classification from histopathological images using patch-based deep learning modeling. IEEE Access, 9: 24273-24287. https://doi.org/10.1109/ACCESS.2021.3056516

[8] Gumaei, A., Sammouda, R., Al-Rakhami, M., AlSalman, H., El-Zaart, A. (2021). Feature selection with ensemble learning for prostate cancer diagnosis from microarray gene expression. Health Informatics Journal, 27(1): 1460458221989402. https://doi.org/10.1177/1460458221989402

[9] Qadri, S.F., Shen, L., Ahmad, M., Qadri, S., Zareen, S.S., Khan, S. (2021). OP-convNet: A patch classification-based framework for CT vertebrae segmentation. IEEE Access, 9: 158227-158240. https://doi.org/10.1109/ACCESS.2021.3131216

[10] Meng, L., Tian, Y., Bu, S. (2020). Liver tumor segmentation based on 3D convolutional neural network with dual scale. Journal of Applied Clinical Medical Physics, 21(1): 144-157. https://doi.org/10.1002/acm2.12784

[11] Fang, X., Xu, S., Wood, B.J., Yan, P. (2020). Deep learning-based liver segmentation for fusion-guided intervention. International Journal of Computer Assisted Radiology and Surgery, 15: 963-972. https://doi.org/10.1007/s11548-020-02147-6

[12] Reyad, M., Sarhan, A.M., Arafa, M. (2024). Architecture optimization for hybrid deep residual networks in liver tumor segmentation using a GA. International Journal of Computational Intelligence Systems, 17(1): 209. https://doi.org/10.1007/s44196-024-00542-4

[13] Qadri, S.F., Shen, L., Ahmad, M., Qadri, S., Zareen, S.S., Akbar, M.A. (2022). SVseg: Stacked sparse autoencoder-based patch classification modeling for vertebrae segmentation. Mathematics, 10(5): 796. https://doi.org/10.3390/math10050796

[14] Kalpana, P., Anandan, R., Hussien, A.G., Migdady, H., Abualigah, L. (2024). Plant disease recognition using residual convolutional enlightened swin transformer networks. Scientific Reports, 14(1): 8660. https://doi.org/10.1038/s41598-024-56393-8

[15] Kalpana, P., Anandan, R. (2023). A capsule attention network for plant disease classification. Traitement du Signal, 40(5): 2051-2062. https://doi.org/10.18280/ts.400523

[16] Abdelwahab, O., Awad, N., Elserafy, M., Badr, E. (2022). A feature selection-based framework to identify biomarkers for cancer diagnosis: A focus on lung adenocarcinoma. Plos ONE, 17(9): e0269126. https://doi.org/10.1371/journal.pone.0269126

[17] Kalpana, P., Chanti, Y., Ravi, G., Regan, D., Pareek, P.K. (2023). SE-Resnet152 model: Early corn leaf disease identification and classification using feature based transfer learning technique. In 2023 International Conference on Evolutionary Algorithms and Soft Computing Techniques (EASCT), Bengaluru, India, pp. 1-6. https://doi.org/10.1109/EASCT59475.2023.10392328

[18] Nabi, S.A., Kalpana, P., Chandra, N.S., Smitha, L., Naresh, K., Ezugwu, A.E., Abualigah, L. (2024). Distributed private preserving learning based chaotic encryption framework for cognitive healthcare IoT systems. Informatics in Medicine Unlocked, 49: 101547. https://doi.org/10.1016/j.imu.2024.101547

[19] Alirr, O.I. (2020). Deep learning and level set approach for liver and tumor segmentation from CT scans. Journal of Applied Clinical Medical Physics, 21(10): 200-209. https://doi.org/10.1002/acm2.13003

[20] Feng, S.L., Zhao, H.M., Shi, F., Cheng, X.N., Wang, M., Ma, Y.H. (2020). CPFNet: Context pyramid fusion network for medical image segmentation. IEEE Transactions on Medical Imaging, 39(10): 3008-3018. https://doi.org/10.1109/TMI.2020.2983721

[21] Rahman, H., Bukht, T.F.N., Imran, A., Tariq, J., Tu, S., Alzahrani, A. (2022). A deep learning approach for liver and tumor segmentation in CT images using ResUNet. Bioengineering, 9(8): 368. https://doi.org/10.3390/bioengineering9080368

[22] Wang, W., Chen, C., Meng, D., Hong, Y., Sen, Z., Jiangyun, L. (2021). Transbts: Multimodal brain tumor segmentation using transformer. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 109-119. https://doi.org/10.1007/978-3-030-87193-2_11

[23] Shao, J., Luan, S., Ding, Y., Xue, X., Zhu, B., Wei, W. (2024). Attention connect network for liver tumor segmentation from CT and MRI images. Technology in Cancer Research & Treatment, 23. https://doi.org/10.1177/15330338231219366

[24] Zhang, Y., Pan, X., Li, C., Wu, T. (2020). 3D liver and tumor segmentation with CNNs based on region and distance metrics. Applied Sciences, 10(11): 3794. https://doi.org/10.3390/app10113794

[25] Wu, J., Zhang, Y., Tang, X. (2019). A joint 3D+2D fully convolutional framework for subcortical segmentation. In Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, pp. 301-309. https://doi.org/10.1007/978-3-030-32248-9_34

[26] Manjunath, R.V., Gowda, Y. (2024). Automated segmentation of liver tumors from computed tomographic scans. Journal of Liver Transplantation, 15: 100232. https://doi.org/10.1016/j.liver.2024.100232

[27] 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: 24898-24909. https://doi.org/10.1109/ACCESS.2019.2899608

[28] 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): 1316-1325. https://doi.org/10.1109/TMI.2019.2948320

[29] Yu, H.J., Son, C.H. (2020). Leaf spot attention network for apple leaf disease identification. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW): Seattle, WA, USA, pp. 229-237. https://doi.org/10.1109/CVPRW50498.2020.00034

[30] Hung, A.L.Y., Zheng, H., Miao, Q., Raman, S.S., Terzopoulos, D., Sung, K. (2022). CAT-Net: A cross-slice attention transformer model for prostate zonal segmentation in MRI. IEEE Transactions on Medical Imaging, 42(1): 291-303. https://doi.org/10.1109/TMI.2022.3211764

[31] Liu, H., Zhuang, Y., Song, E., Xu, X., Ma, G., Cetinkaya, C., Hung, C.C. (2023). A modality-collaborative convolution and transformer hybrid network for unpaired multi-modal medical image segmentation with limited annotations. Medical Physics, 50(9): 5460-5478. https://doi.org/10.1002/mp.16338

[32] Almotairi, S., Kareem, G., Aouf, M., Almutairi, B., Salem, M.A.M. (2020). Liver tumor segmentation in CT scans using modified SegNet. Sensors, 20(5): 1516. https://doi.org/10.3390/s20051516

[33] Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M. (2022). Swin-unet: Unet-like pure transformer for medical image segmentation. In Computer Vision-ECCV 2022 Workshops, Tel Aviv, Israel, pp. 205-218. https://doi.org/10.1007/978-3-031-25066-8_9

[34] Liver Tumor Segmentation. https://www.kaggle.com/datasets/andrewmvd/liver-tumor-segmentation.