Computers in Biology and Medicine 179 (2024) 108870

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

A dual-encoder double concatenation Y-shape network for precise volumetric liver and lesion segmentation

Gabriella d'Albenzio [a,b,*,1], Yuliia Kamkova [d,b,1], Rabia Naseem [c], Mohib Ullah [f], Stefania Colonnese [e], Faouzi Alaya Cheikh [f], Rahul Prasanna Kumar [a]

[a] The Intervention Center, Oslo University Hospital, Oslo, Norway
[b] Department of Informatics, University of Oslo, Oslo, Norway
[c] COMSATS, University Islamabad, Islamabad, Pakistan
[d] Department of Research and Development, Division of Emergencies and Critical Care, Oslo University Hospital, Oslo, Norway
[e] Department of Information Engineering, Electronics and Telecommunications (DIET), La Sapienza University of Rome, Rome, Italy
[f] Department of Computer Science, Norwegian University of Science and Technology, Gjøvik, Norway

ARTICLE INFO

Keywords:
CNN
Abdominal CT
Multi-class segmentation
HCC
Liver

ABSTRACT

Accurate segmentation of the liver and tumors from CT volumes is crucial for hepatocellular carcinoma diagnosis and pre-operative resection planning. Despite advances in deep learning-based methods for abdominal CT images, fully-automated segmentation remains challenging due to class imbalance and structural variations, often requiring cascaded approaches that incur significant computational costs. In this paper, we present the Dual-Encoder Double Concatenation Network (DEDC-Net) for simultaneous segmentation of the liver and its tumors. DEDC-Net leverages both residual and skip connections to enhance feature reuse and optimize performance in liver and tumor segmentation tasks. Extensive qualitative and quantitative experiments on the LiTS dataset demonstrate that DEDC-Net outperforms existing state-of-the-art liver segmentation methods. An ablation study was conducted to evaluate different encoder backbones — specifically VGG19 and ResNet — and the impact of incorporating an attention mechanism. Our results indicate that DEDC-Net, without any additional attention gates, achieves a superior mean Dice Score (DS) of **0.898** for liver segmentation. Moreover, integrating residual connections into one encoder yielded the highest DS for tumor segmentation tasks. The robustness of our proposed network was further validated on two additional, unseen CT datasets: IDCARDb-01 and COMET. Our model demonstrated superior lesion segmentation capabilities, particularly on IRCADb-01, achieving a DS of 0.629. The code implementation is publicly available at this website.

1. Introduction

According to the Global Cancer Observatory [1], liver cancer is the sixth most frequently diagnosed cancer and the third leading cause of cancer death worldwide. Early diagnosis and treatment can significantly reduce mortality rates. Computer-assisted systems for diagnosis, surgery planning, and navigation have been instrumental in aiding radiologists and surgeons, thereby improving the efficiency and accuracy of surgical interventions [2]. Segmentation is crucial in the pre-operative planning workflow, as it assigns unique labels to different anatomical and pathological structures in medical images [2,3]. While many deep-learning methods have been proposed for medical image segmentation, fully-automated segmentation of the liver and its lesions still presents a significant challenge [4,5]. Automated liver and tumor segmentation is a complex process, hampered by numerous obstacles. The initial step in any liver resection procedure is the segmentation of the liver [6]. However, due to the liver's soft-tissue nature and its location in close proximity to adjacent organs such as the spleen, stomach, and gall bladder, identifying its boundaries can be challenging. Additionally, the presence of air or gas in the gastrointestinal tract can hinder the visualization of these organs on CT scans. Furthermore, the appearance and form of the liver can be altered by injuries and other pathologies, leading to changes in its thickness, signal intensity, and overall architecture. Lesion segmentation further complicates the process due to low contrast in CT images, varying tumor shapes, and significant class imbalance, where lesions are much smaller than the liver. These challenges can affect the accuracy of automated segmentation processes and require careful consideration during liver resection procedures. State-of-the-art segmentation approaches, such as UNet [7] and its variants, often fail

[*] Corresponding author at: The Intervention Center, Oslo University Hospital, 0slo, Norway.
E-mail address: gabridal@uio.no (G. d'Albenzio).
[1] Authors contributed equally.

https://doi.org/10.1016/j.compbiomed.2024.108870
Received 26 August 2023; Received in revised form 7 July 2024; Accepted 8 July 2024
Available online 17 July 2024

with imbalanced data because high-resolution information lost in the contracting path cannot be fully recovered in the expansion path. In contrast, Y-shaped networks with two encoders and one decoder have gained popularity for their flexibility in capturing deeper and more diverse features. To address these challenges, we propose a novel dual-encoder architecture that leverages the complementary strengths of two different encoders, enhancing feature representation and robustness. The dual-encoder structure allows the network to learn from two perspectives: one encoder (VGG-19) focuses on capturing detailed spatial features, while the other (ResNet) emphasizes deeper contextual information through residual connections. The proposed approach leverages contextual volumetric information in 3D medical imaging and employs skip connections to combine deep coarse semantic information with fine-grained details, enhancing the model's generalization capability. Additionally, we explore the benefits of incorporating an Attention Gate (AG) mechanism to refine the segmentation by emphasizing relevant contextual information from CT scans. The key contributions of our work are outlined as follows:

1. We propose a novel end-to-end architecture, with dual-branch encoders and one decoder for automated liver and tumour segmentations in CT scans. This model effectively employs a double concatenation of feature maps from lower to higher resolution levels, significantly enhancing spatial information capture.
2. To extract diverse features and create a streamlined end-to-end network, our method adopts a heterogeneous architecture. It consists of two distinct encoders: one leveraging the robust feature extraction capabilities of VGG-19, and the other utilizing residual connections to facilitate a more efficient, deeper model. This dual-encoder system is integrated with a single decoder, maintaining the encoder–decoder structure and skip connections characteristic of the original UNet.
3. We have demonstrated that the proposed Y-shaped network achieves high segmentation accuracy for both liver and lesions without the need for an additional attention mechanism. This highlights the inherent effectiveness and robustness of our network architecture in capturing and segmenting relevant features accurately.
4. We conduct an extensive re-implementation of existing state-of-the-art (SOTA) methods on the LiTS2017, and validate the robustness of our proposed network on 3DIRCADb and OSLO-COMET datasets. Our experimental results confirm the superiority of our approach across multiple metrics.

2. Related work

Medical image segmentation from 3D volumetric images, such as computed tomography (CT) scans and magnetic resonance imaging (MRI), has been a key focus in medical image processing for decades due to its ability to preserve the 3D structure and spatial information of internal body structures. Deep learning, in particular, has shown significant advancements and superior performance in this area [8], as it can learn hierarchical feature representations at multiple levels of abstraction.

Fully Convolutional Neural Networks (FCNNs) have been highly effective in medical image segmentation, excelling in delineating not only the liver but also many other organs and structures [9,10]. The choice between 2D and 3D FCNNs depends on the task specifics and available computational resources. Unlike typical Convolutional Neural Networks (CNNs) that incorporate fully connected layers at the end of the architecture, FCNNs consist solely of convolutional and, optionally, pooling layers. This design eliminates the high parameter count associated with fully connected layers, enabling FCNNs to learn spatial hierarchies of features through backpropagation [11]. Hu et al. [12] developed a deep 3D CNN that predicts a probability map as a subject-specific prior, assigning each voxel the likelihood of being liver tissue. Among FCN-based architectures, UNet stands out as a baseline standard for medical image segmentation, available in both 2D [7] and 3D versions [13]. UNet uses skip connections to combine information from bottom and top layers, addressing the challenge of recovering original data during upsampling and mitigating information loss. A notable cascade 2D FCNN architecture, involving two UNets, was proposed by Christ et al. They train and cascade two FCNs: the first segments the liver, and the second uses the predicted liver ROI to segment lesions. The authors also employ dense 3D conditional random fields for refining segmentation results, accounting for spatial coherence and appearance [4]. Furthermore, Milletari et al. [14] introduced a symmetric end-to-end 3D FCNN with residual connections and a novel loss function to better handle class imbalance than the classic weighted cross-entropy loss. Subsequently, the Hybrid Densely Connected UNet (H-DenseUNet) was proposed, integrating 3D DenseUNet with 2D DenseUNet using auto-context for liver and tumor segmentation [15].

To simplify the complexity of cascaded approaches, Xu et al. introduced ResUNet for liver segmentation. This model enhances the UNet by adding residual units and batch normalization layers to both the upsampling and downsampling parts, resulting in a deeper network that achieves quick convergence [16]. Zhou et al. proposed UNet++, a more flexible architecture with nested skip connections designed to bridge the semantic gap between encoder and decoder feature maps, improving gradient flow [17]. UNet3+ [18] advanced dense skip connections by employing full-scale skip connections, though it can struggle with segmenting small objects when training data is limited. UNet# [19] combined dense-scale and full-scale connections between the encoder and decoder subnets, enhancing the network's ability to explore full-scale information.

To enhance the performance of UNet, many researchers have focused on the encoder–decoder architecture, introducing novel mechanisms to improve segmentation outcomes. Among these, the *attention* mechanism is particularly notable. Inspired by cognitive attention, it selectively processes relevant stimuli [20]. Initially applied to natural language processing tasks, attention mechanisms have recently shown great promise in computer vision and medical segmentation tasks, integrating various types of attention mechanisms into the encoder–decoder structure. Research has shown that incorporating attention gates into deep learning architectures improves network performance [21–23]. Inspired by this, Oktay et al. [21] introduced an attention gate (AG) model to focus on structures of different shapes and sizes, filtering out unnecessary information and enhancing the network's ability to focus on relevant features. Schlemper et al. [24] utilized a self-gated soft attention mechanism that generates an end-to-end trainable gating signal, contextualizing local information for more accurate predictions. The Attention Hybrid Connection Network (AHCNet) [25] combines a cascaded 3D FCNN with densely connected long and short skip connections, and hard and soft attention, to segment liver and tumors. This work introduced a joint Dice loss function to handle class imbalance and a focal loss to reduce false positives, concluding that hybrid connections and attention mechanisms improve training speed and accuracy. RA-UNet leverages the strengths of UNet, residual learning, and the attention residual mechanism to focus on relevant image parts and suppress irrelevant ones [26]. Recently, the Residual Multi-scale Attention UNet (RMAU-Net) [27] was introduced for liver and tumor segmentation, combining residual attention learning and squeeze-excitation blocks with UNet. This model has shown promising results in segmenting both liver and tumors in CT volumes.

However, the U-shaped networks mentioned above often struggle with achieving accurate results near organ or tumor boundaries. To address this, a new family of networks, Y-shaped networks, has shown promising results in various medical image segmentation tasks [28–30]. These networks utilize a dual-branch encoder structure to simultaneously consider both global and local context, reducing the impact of organ/tumor location and shape variability.

Our objective is to investigate whether integrating residual connections within the encoder of a Y-shaped architecture, along with skip connections between the encoder and decoder, can improve performance. Additionally, we explore the hypothesis that the core architecture of the encoder network may not be as crucial as the presence of skip connections, suggesting effective segmentation can be achieved even without a sophisticated encoder. This exploration seeks to understand the impact of these architectural components on segmentation accuracy, aiming for advancements in medical imaging analysis.

3. Method

In this section, we present an overview of the core components of our proposed DEDC-Net, a Dual-Encoder Double Concatenation Network for liver and tumor segmentations. The network architecture consists of two encoders (VGG19 [31] and ResNet [32]) and one decoder similar to 3D UNet [7] that are connected through skip connections. The input to the network is a computed tomography (CT) volume $x \in \mathbb{R}^{D\times H\times W\times 1}$, and its corresponding multi-class segmentation label $y \in \mathbb{Z}^{D\times H\times W\times C}$, where D, H, W represent the dimensions of the voxels and C is the number of classes (0-background, 1-liver, 2-tumor, respectively). The output of the network is a predicted segmentation map $\hat{y}$. Furthermore, we also present a brief description of the attention mechanism, as we conducted experiments integrating the attention gate components proposed in [21] into our 3DY-Net architecture and other competitive networks.

3.1. DEDC-Net architecture

This study introduces a novel architecture for volumetric segmentation with a Y-shaped structure. The proposed network, DEDC-Net, builds upon previous Y-shaped neural networks used for tasks such as polyp detection [33], retinal OCT segmentation [28], and breast biopsy image diagnosis [34]. The key innovation of our work is the integration and double concatenation of features from the two encoders at each stage of the decoder, as shown in Fig. 1. To our knowledge, this is the first application of a Y-shaped network for volumetric segmentation of the liver and its lesions.

Encoders: We experimented with two encoder configurations: a modified 3D VGG19 [31] and a modified 3D ResNet [32], both excluding the three fully connected layers. For the first encoder branch, we used VGG-19, known for its 19 deep layers. The initial two layers consist of two 3 × 3 × 3 convolutions with stride and padding of 1, each followed by a rectified linear unit (ReLU) and a 2 × 2 × 2 max pooling operation with strides of 2. The subsequent three layers follow a similar pattern but include four convolutions each. For the second branch, we chose a modified ResNet [32], incorporating five layers with residual connections to mitigate the vanishing gradient problem. This configuration includes 27 convolutional blocks, with the first layer using 7 × 7 × 7 filters and the rest using 3 × 3 × 3 filters. Only two pooling layers are employed—one at the beginning and one at the end.

Decoder: Our network's decoder is similar to the original 3D UNet [13] one. U-Net-shaped models are classical fully-convolutional neural networks for classification, segmentation, and detection in medical imaging, consisting of a contracting and an expanding path. The decoder has four convolutional blocks, and the feature maps are upsampled by using a 3 × 3 × 3 transposed convolution operation with strides of 2 in each dimension, followed by three 3 × 3 × 3 convolutions, each followed by a SELU [35]. Finally, a 1 × 1 × 1 convolution operation generates the final segmentation map.

Detailed descriptions of the double concatenations between the encoders and the decoder are provided in the following subsection to ensure a comprehensive understanding of the network's design.

3.2. Double concatenation formulation for dual-encoder volumetric segmentation network

In our proposed network architecture for volumetric segmentation of liver and tumors in medical images, we employ a dual-encoder setup consisting of one 3D VGG encoder and one 3D ResNet encoder, followed by a single decoder. The integration of outputs from both encoders into the decoder is achieved through a novel double concatenation process, which ensures the effective combination of feature maps. This process is described as follows:

Let E_{VGG} and E_{ResNet} represent the feature maps from the 3D VGG encoder and the 3D ResNet encoder, respectively. The double concatenation process is executed in the following stages:

Initial concatenation. The initial feature maps from both encoders are concatenated to form a combined feature map C_{center}:

$$C_{\text{center}} = \text{Concat}(E_{VGG,0}, E_{ResNet,0})$$

This combined feature map is then processed through a series of convolutional blocks:

$$C_{\text{center}} = \text{ConvBlock}_{c0}(\text{ConvBlock}_{c1}(C_{\text{center}}))$$

Concatenation at each decoder stage. For each subsequent stage i in the decoder, the feature maps from each encoder are concatenated with the upsampled feature maps from the previous decoder layer f_i:

$$f_i = \text{Upsample}_i(C_{\text{center}})$$

$$C_{VGG,i} = \text{Concat}(E_{VGG,i}, f_i)$$

$$C_{ResNet,i} = \text{Concat}(E_{ResNet,i}, f_i)$$

Final concatenation. The outputs of these initial concatenations are then concatenated again to create the final feature map C_{final} :

$$C_{\text{final},i} = \text{Concat}(C_{VGG,i}, C_{ResNet,i})$$

The final concatenated feature map is passed through additional convolutional blocks and then serves as the input for the next layer of the decoder, resulting in D_i:

$$D_i = \text{ConvBlock}_{i,0}(\text{ConvBlock}_{i,1}(\text{ConvBlock}_{i,2}(C_{\text{final},i})))$$

This double concatenation approach effectively leverages the complementary strengths of the 3D VGG and 3D ResNet encoders to enhance feature representation.

3.3. Attention gate component

In our networks, the AGs are incorporated into the decoder stage of the architecture through each skip connection. As shown in Fig. 2, a gating signal vector g is collected from a coarser scale as it encodes global information from a large spatial context. Let be $F^l_{x_i}$ the input feature maps of a layer l and $F^{l-1}_{g_i}$ the feature map of the gating vector g_i used for each voxel i, collected from a coarser scale and with encoded global information from a large spatial context. The AG module extracts a 3D attention coefficients α_i, which can be mathematically represented as:

$$\alpha_i = \sigma_2[\Phi(\sigma_1(W_x F^l_{x_i} + W_g F_{g_i}))]; \tag{1}$$

where $\sigma_1(x)$ is an element-wise nonlinearity (i.e. rectified linear-unit ReLU), $\sigma_2(x)$ is a normalization function (i.e Sigmoid function), W_x, W_g are convolutional operations and Φ is the upsampling function. In this way, the output of AGs is the element-wise multiplication $F^l_{x_i}$ of input feature maps $F^l_{x_i^{out}}$ and attention coefficients α_i:

$$F^l_{x_i^{out}} = \alpha_i \times F^l_{x_i} \tag{2}$$

Specifically, in DEDC-Net, the attention gate is added before each upsampling operation of each block and receives feature maps from the current and the corresponding downsampling block, as shown in Fig. 3

3.4. Multi-class loss function

Our networks are trained with a combination of Soft Dice Loss and Cross-Entropy loss. The detailed description of both loss functions can be found here [36]. Let $\hat{Y}$ be the reference foreground segmentation

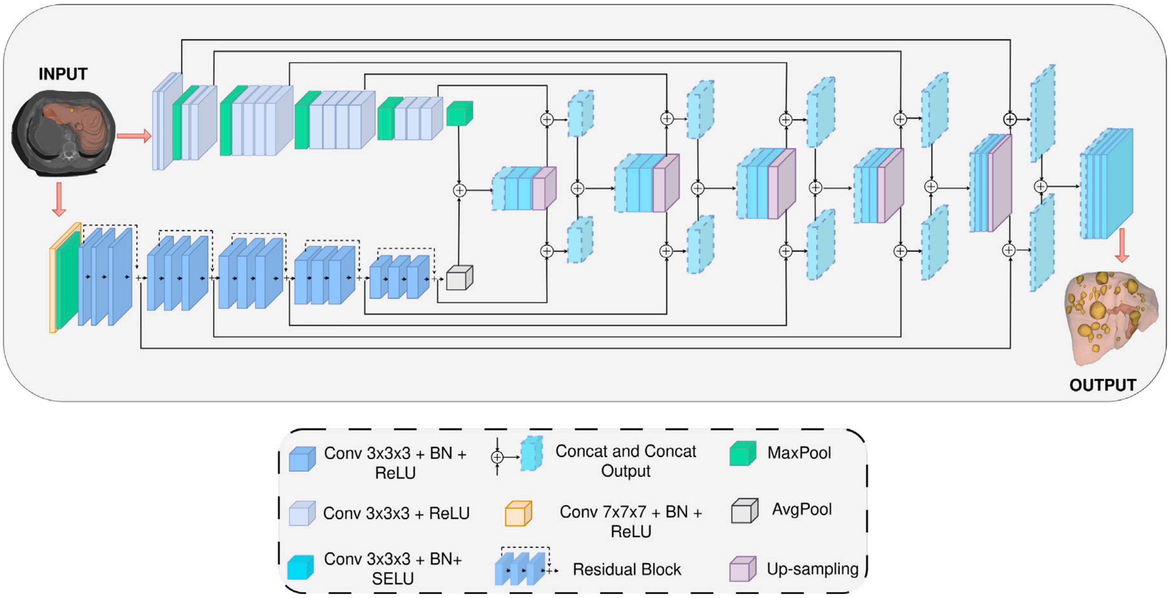

Fig. 1. Architecture of the Dual-Encoder Double Concatenation Network (DEDC-Net) for volumetric liver and tumor segmentation. The network consists of dual-encoder branches: one 3D VGG encoder (top) and one 3D ResNet encoder (bottom), followed by a single decoder. The feature maps from both encoders are concatenated at each stage using a double concatenation approach.

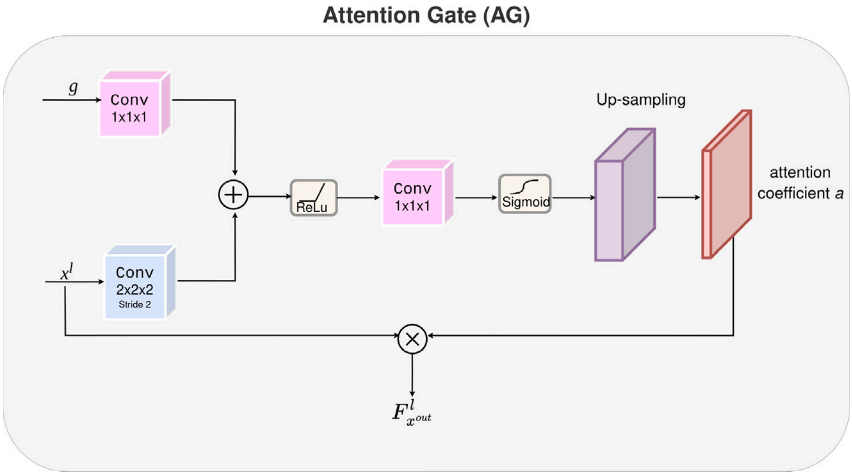

Fig. 2. Visual representation of the 3D Attention Gate (AG) mechanism, which selectively focuses on important features within a three-dimensional space.

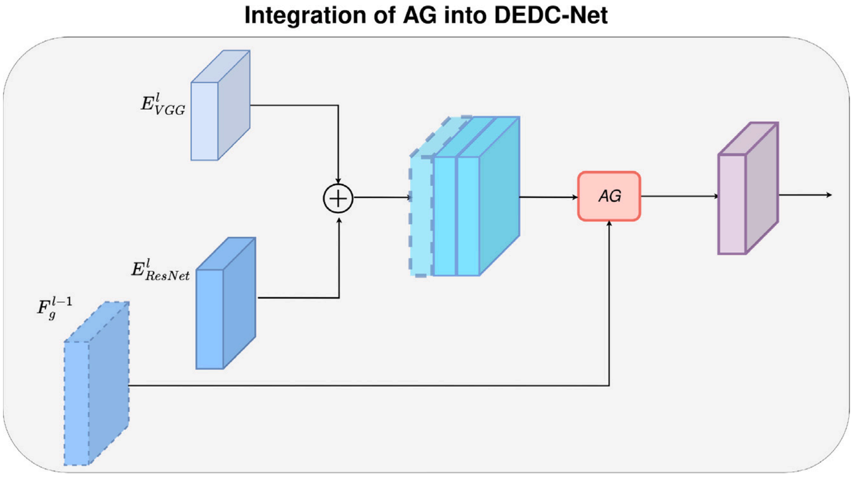

Fig. 3. Integration of the Attention gate (AG): E^l_{VGG} and E^l_{ResNet} represent feature maps from the two encoder paths, and the F_g^{l-1} the feature map of the gating vector from the previous convolutional layer. For details on the core components, please refer to the decoder path illustrated in Fig. 1.

(ground truth) with voxel values $\{\hat{y}\}_n = 1, \ldots, N$, and P the predicted probabilistic map for the foreground label over N image elements $\{p_n\}_n = 1, \ldots, N$, with the background class probability being 1-P. The ϵ provides numerical stability to prevent division by zero, and $\{c_n\}_n = 1, \ldots, C$ indicates the class label. The Soft Dice Loss is defined as:

$$\mathcal{L}_{DL} = 1 - \frac{\sum_{n=1}^{N} \sum_{c=1}^{C} p_n^c \hat{y}_n^c + \epsilon}{\sum_{n=1}^{N} \sum_{c=1}^{C} p_n^c + \hat{y}_n^c + \epsilon} \tag{3}$$

and the Cross-Entropy Loss is as follows:

$$\mathcal{L}_{CE} = -\frac{1}{N} \sum_{n=1}^{N} \sum_{c=1}^{C} r_n \lg(p_n^c) + (1 - \hat{y}_n^c) log(1 - p_n) \tag{4}$$

Then, the combined loss function [37] is expressed as:

$$\mathcal{L}_T = \mathcal{L}_{DL} + \mathcal{L}_{CE}; \tag{5}$$

3.5. Evaluation metrics

To quantitatively assess segmentation performance, we employ various metrics:

1. Average Symmetric Surface Distance (ASSD) measures the average distance between points on the segmented and ground truth surfaces:

   $$\mathrm{ASSD}(Pr, GT) = \frac{1}{|S(Pr)| + |S(GT)|} \left(\sum_{s_{Pr} \in S(Pr)} d(s_{Pr}, S(GT)) + \sum_{s_{GT} \in S(GT)} d(s_{GT}, S(Pr)) \right) \tag{6}$$

   where $S(Pr)$ and $S(GT)$ represent the surfaces of predicted and ground truth objects, $|S(Pr)|$ and $|S(GT)|$ denote their cardinalities, and $d(s_{Pr}, S(GT))$ and $d(s_{GT}, S(Pr))$ are distances from points on the surfaces.
2. Volume Similarity (VS) quantifies the similarity between segmented and ground truth volumes:

   $$\mathrm{VS}(Pr, GT) = 1 - \frac{||Pr| - |GT||}{|Pr| + |GT|} \tag{7}$$

   where $|Pr|$ and $|GT|$ are volumes of segmented and ground truth objects, respectively.
3. SN (Sensitivity) measures the ratio of true positive predictions to all positive instances:

   $$\mathrm{SN} = \frac{TP}{TP + FN} \tag{8}$$

4. Specificity (SP) calculates the ratio of true negative predictions to all negative instances:

   $$\mathrm{SP} = \frac{TN}{TN + FP} \tag{9}$$

5. Precision (PR) computes the ratio of true positive predictions to all positive predictions:

   $$\mathrm{PR} = \frac{TP}{TP + FP} \tag{10}$$

6. Dice Score (DS) measures the overlap between predicted and ground truth segmentation masks:

   $$\mathrm{DS} = \frac{2 \times |Pr \cap GT|}{|Pr| + |GT|} \tag{11}$$

4. Experiments setup

4.1. Datasets

4.1.1. LiTS

The LiTS dataset, originating from the Liver Tumor Segmentation Challenge hosted by ISBI 2017 and MICCAI 2017, stands as the preeminent publicly available dataset for conducting studies on liver and tumor segmentation. Comprising 201 contrast-enhanced abdominal CT scans, the dataset includes 131 scans annotated for training purposes and 70 designated for testing, sourced from six distinct clinical sites utilizing various scanners and protocols. This dataset is characterized by its wide variation in spatial resolution and field of view, with in-plane resolutions ranging between 0.60 mm to 0.98 mm and slice spacings from 0.45 mm to 5.0 mm. Each scan features axial slices of uniform size (512 × 512 pixels), though the count of slices along the z-axis varies significantly, ranging from 42 to 1026. The large amount of CT volumes and segmentations make the LiTS dataset a valuable resource for deep learning applications in liver and tumor segmentation.

4.1.2. 3D-IRCADb-01

The 3D Image Reconstruction for Comparison of Algorithm Database (3D-IRCADb) is commonly used in medical imaging research. It is publicly available and widely utilized by researchers in this field. These images contain patient data that has been anonymized, along with manually delineated regions of interest by medical professionals. The 3D-IRCADb-01 subset consists of enhanced CT scans from 20 individuals, with an equal distribution between women and men. Among the female cohort, 75% of them have liver tumors. The dataset has a resolution of 512 × 512 pixels, with the z-axis containing 91 to 260 slices.

4.1.3. OSLO-COMET

The clinical dataset used in the paper is the subset of 15 abdominal CT scans with liver and colorectal metastasis from the ethnically approved Oslo-CoMet Study (COMET) [38]. The CT imaging was acquired at a resolution of 512 × 512 pixels, with each volume comprising 94 to 427 slices. The selection criteria for this subset were rigorously defined by the position and size of liver lesions, specifically targeting lesions with diameters ranging from 3 mm to 10 mm.

4.2. Implementation details

We train, validate and test our network on the widely-used LiTS dataset of MICCAI 2017 Liver Tumor Segmentation Challenge [39] and evaluate the robustness of achieved results on 3D-IRCADb-01 and COMET datasets. The division of the LiTS dataset into training, validation, and testing subsets is executed randomly, with the selection based only on the available training set. This approach is necessitated due to the absence of publicly accessible labels for the test set. Consequently, the dataset allocation consists of 101 scans for training, 15 for validation, and 15 for testing purposes. For our experiments, all images are resized to 256 × 256 × 64 due to the GPU limitation. Moreover, since the focus of this task is the liver and lesions, the Hounsfield unit (HU) values were windowed in the range [−175], [250] to exclude artifacts and irrelevant organs and tissues. For intensifying the network generalization, data augmentation is done "on-the-fly" during training, including a series of geometric transformations, such as random flipping, shifting, scaling, and resampling. The implementation of our network is based on PyTorch 1.12.0 [40]. The network is trained for 500 epochs on an NVIDIA GeForce RTX 3090 graphic memory with a batch size of 1. The best validation accuracy for all models was used to determine the number of training epochs. The normal distribution initializer [41] is employed for initializing the weights since its robustness in considering the rectifier nonlinearities. For training our deep neural network, the learning rate adaptive optimizer ADAM [42], was used. ADAM optimizer dominates the field of deep learning due to its fast convergence. We set the initial learning rate to 0.0001, decaying the learning rate with a cosine annealing for each batch as proposed in [43]. For the network robustness evaluation, we took the network with the best performance on the test LITS subset and trained it on the full dataset for another 100 epochs. Then the inference results were inspected with the reference to evaluation metrics.

Table 1
Quantitative comparison of different methods on the LiTS dataset (15 test volumes). Red indicates the best scores for liver segmentation, while blue indicates the best scores for lesion segmentation.

Network	Class	ASSD	VS	SN	SP	PR	DS
3DUNet	Liver	**0.583 ± 0.373**	0.928 ± 0.031	**0.936 ± 0.094**	0.995 ± 0.002	0.847 ± 0.051	0.885 ± 0.049[a]
3DUNet+AG		0.823 ± 26.930	0.919 ± 0.270	0.922 ± 0.123	0.994 ± 0.002	0.837 ± 0.041	0.871 ± 0.059
3DVNet		1.443 ± 1.815	0.908 ± 0.064	0.880 ± 0.145	0.996 ± 0.002	0.845 ± 0.059	0.852 ± 0.083
3DVNet+AG		2.967 ± 4.014	0.649 ± .0164	0.816 ± 0.251	0.996 ± 0.003	**0.855 ± 0.083**	0.806 ± 0.166[b]
3DResUNet		1.271 ± 1.180	0.917 ± 0.049	0.926 ± 0.072	0.993 ± 0.004	0.795 ± 0.066	0.853 ± 0.50
UNETR		1.415 ± 1.556	0.917 ± 0.052	0.937 ± 0.032	0.993 ± 0.032	0.797 ± 0.074	0.859 ± 0.046
SwinTr		1.421 ± 1.931	**0.956 ± 0.045**	0.921 ± 0.047	0.956 ± 0.003	0.849 ± 0.074	0.857 ± 0.050
DEDC-Net (Ours)		0.873 ± 0.915	0.918 ± 0.038	0.935 ± 0.085	**0.993 ± 0.003**	0.806 ± 0.063	**0.898 ± 0.031**
3DUNet	Lesions	25.676 ± 36.177	0.645 ± 0.297	0.400 ± 0.386	0.999 ± 0.002	0.390 ± 0.321	0.384 ± 0.350
3DUNet+AG		24.249 ± 35.612	0.635 ± 0.319	0.430 ± 0.357	0.996 ± 0.001	0.400 ± 0.285	0.380 ± 0.296
3DVNet		13.297 ± 19.022	0.564 ± 0.339	0.370 ± 0.336	0.999 ± 0.001	**0.500 ± 0.376**	0.351 ± 0.302
3DVNet+AG		22.127 ± 22.433	0.484 ± 0.294	0.360 ± 0.335	0.998 ± 0.003	0.360 ± 0.345	0.268 ± 0.262[b]
3DResUNet		**10.535 ± 10.499**	0.44 ± 0.330	0.337 ± 0.348	0.999 ± 0.001	0.45 ± 0.367	0.325 ± 0.280[a]
UNETR		20.454 ± 19.377	0.538 ± 0.313	0.287 ± 0.262	0.999 ± 0.001	0.314 ± 0.323	0.267 ± 0.258[b]
SwinTr		22.343 ± 19.113	0.538 ± 0.347	0.343 ± 0.319	0.998 ± 0.001	0.261 ± 0.282	0.252 ± 0.256[b]
DEDC-Net (Ours)		12.168 ± 12.666	**0.673 ± 0.282**	**0.570 ± 0.323**	0.998 ± 0.003	0.464 ± 0.310	**0.461 ± 0.274**

[a] Denotes $p < 0.05$.

[b] Denotes $p < 0.01$.

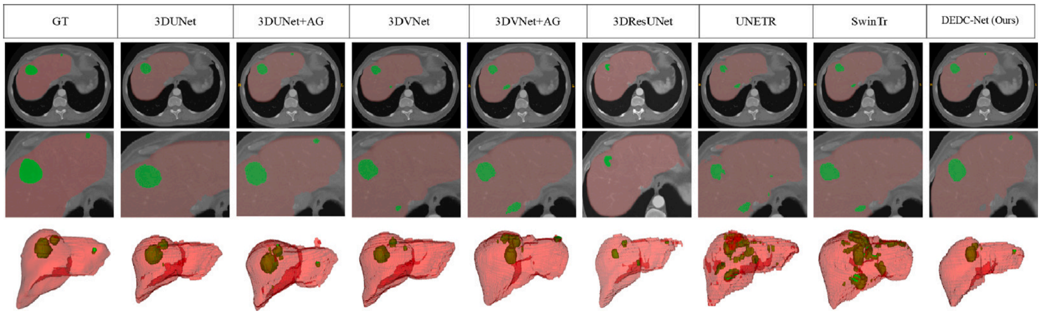

Fig. 4. Visual assessment of the competing networks for liver and tumors segmentation on the volume 072 of LiTS dataset: the first two rows represent the axial view of CT images of a patient from the test set. The third row illustrates the corresponding 3D models generated from segmentations. GT: ground truth. Red: liver segmentation. Green: tumor segmentation.

5. Experimental results

5.1. Comparison of models

To validate the effectiveness of the proposed DECD-Net, we compared it with leading segmentation models commonly used in medical image segmentation. We utilized the 3D versions of various networks, including 3DUNet [13], 3D VNet [44] with and without the AGs [21, 45], and the 3D version of ResUNet [46], which integrates a residual network with UNet, theoretically providing deep network layers and robust feature extraction capabilities. Additionally, we compared our network with the latest state-of-the-art models, such as SwinTr [47], which replaces the convolutional layers in UNet with Swin transformer layers, relying entirely on transformer models, and UNETR [48], that employs transformers to capture global information of UNet and has demonstrated high performance in various organ segmentation tasks from CT scans. Employing the same dataset and split selection approach, our objective was to assess segmentation performance.

Compared to other networks outlined in Table 1, our proposed DEDC-Net shows competitive performance across various metrics for liver and tumor segmentations. Specifically, DEDC-Net achieved the highest scores in segmentation DS, SN, SP, and VS. For tumor segmentation, our network outperformed the 3DUNet by nearly 20% in terms of the Dice score. While it did not lead in metrics such as ASSD and PR, the results were still comparable to other leading networks. Lesion segmentation results were relatively low. However, this may be attributed to GPU memory limitations and the need for down-sampling volumes for further processing.

To visualize the results, we present segmentation outcomes for two test cases with significant differences in DS performance: volumes 72 and 130. Figs. 4 and 5 show qualitative results comparing the performance of the competitive models. In Volume 72 (Fig. 4), our proposed network achieved a DS of 0.938 for liver segmentation and 0.789 for lesion segmentation. Visual inspection revealed that only DEDC-Net and 3DUNet+AG detected a small tumor in the left lobe of the liver. Notably, 3DUNet tended to over-segment lesions, while our network demonstrated slight under-segmentation.

In Volume 130, the DS was lower, with 0.894 for liver segmentation and 0.560 for tumor segmentation. None of the networks successfully detected and fully segmented the large tumor in the left lobe of the liver, and identifying multiple small tumors in the right lobe remained challenging for all models. Fig. 5 illustrates that DEDC-Net slightly outperformed other networks in liver parenchyma segmentation, although there was a significant false positive tumor detected in the right lobe. Our network was the only one to segment the large tumor with better performance, but none of the networks successfully detected or segmented the small tumors in the right lobe.

To demonstrate that DEDC-Net improves segmentation accuracy due to its Y-shaped design and double concatenation, rather than adding more parameters, we list the number of parameters for each model in Table 2. Our proposed network has 6.6M more parameters than the 3D-UNet but significantly fewer than the latest state-of-the-art networks like SwinTr or UNETR. Despite having the highest number of parameters (91.1M), UNETR presents one of the lowest scores in lesion segmentation. This results in more efficient training, better generalization, reduced computational cost, quicker inference, and easier optimization for our network.

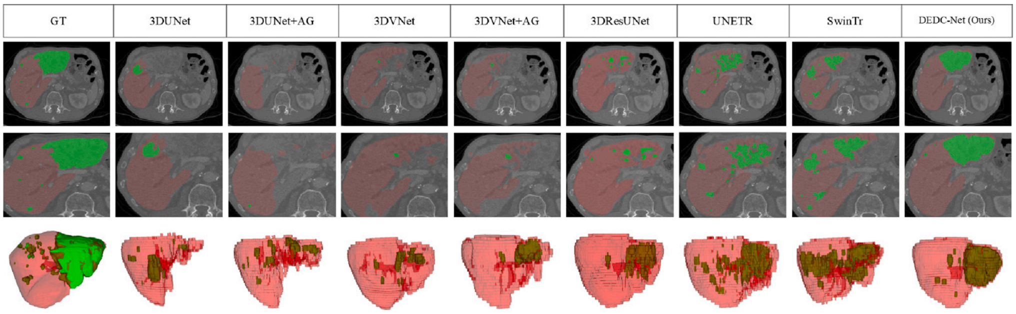

Fig. 5. Visual comparison of the competing networks for liver and tumors segmentation on the volume 130 of LiTS dataset: the first two rows represent the axial view of CT images of a patient from the test set. The third row illustrates the corresponding 3D models generated from segmentations. GT: ground truth. Red: liver segmentation. Green: tumor segmentation.

Table 2
Parameter counts for various models used in the State-of-the-Art (SOTA) analysis. This comparison includes our proposed model, DEDC-Net, alongside other competing networks.

Method	Parameters (M)
3DUNet	19.07
3DUNet+AG	29.11
VNet	45.60
3DVNet+AG	45.58
3DResUNet	65.40
UNETR	91.10
SwinTr	51.93
DEDC-Net (Ours)	25.68

5.2. Ablation analysis

In the first section, we present ablation studies to evaluate the effectiveness of each component within our proposed framework. The variables in the ablation study include the choice of encoder network (VGGNet or ResNet) and the method of using skip connections. The skip connections can be of the Encoder-Decoder type (ED), where the feature map is first concatenated between encoders and decoders, and then concatenated among themselves as described in Section 3.2. Alternatively, the Encoder-Encoder type (EE) is used, where the feature maps from each encoder are first concatenated among themselves and then integrated into the decoder layers of equal resolution, inspired by the approach proposed in [33]. Another area of investigation was the integration of the attention mechanism (AG). Our evaluation focuses on the average ASSD, VS, SN, SP, PR, and Dice metrics across the LiTS test dataset (15 CT volumes).

From the ablation study (Table 3), we can observe the impact of long-range skip connections and the choice of encoder network in our model. Using skip connections directly between the encoder and decoder (ED) generally produces better segmentation results than first using skip connections between both encoders and then concatenating the map to the decoder (EE type). This ED approach yields the highest scores in terms of volumetric similarity, sensitivity, and Dice score for both liver and tumor segmentation.

Integrating the attention mechanism into the framework did not improve segmentation scores. In fact, frameworks without the attention gate (AG) demonstrated higher ASSD, VS, SN, SP, and Dice scores for both liver and tumor segmentation.

Regarding the choice of the network for the feature encoder, ResNet showed improvements in precision scores, while VGGNet enhanced volumetric similarity in both liver and tumor segmentation.

In summary, the combination of two different encoders, such as VGGNet and ResNet, using the ED type of skip connection without the AG, resulted in the best DS for both liver parenchyma and lesion segmentation. This configuration achieved a DS of 0.897 ± 0.030 for the liver and 0.461 ± 0.247 for the lesions.

5.3. Assessment of network generalizability

The proposed DEDC-Net was trained using the complete LITS dataset, where ground truth segmentations were available (130 volumes). Below, we present the results obtained from applying these trained models to the COMET and 3D-IRCADb-01 datasets, as summarized in Table 4.

The segmentation performance results of our proposed network on the unseen 3D-IRCADb-01 and COMET datasets demonstrate promising outcomes in liver and lesion segmentation.

On the 3D-IRCADb-01 dataset, our network achieved a Dice score of 0.629 for tumor segmentation, with an average ASSD of 0.648 and a volumetric similarity of 0.910.

In the COMET dataset, our network showcased robust performance in liver segmentation, consistent with earlier results presented in this paper. The average Dice score for the liver segmentation achieved by network is 0.854, while for lesion segmentation is 0.435. Despite the challenging nature of the task, the segmentation results indicate the network's capability to accurately localize liver parenchyma and detect tumors.

6. Discussion

Automatic liver and tumour segmentation plays a critical role in clinical planning. It can reduce the amount of time clinicians dedicate to this task while aiding in the diagnosis process. In this study, we present DEDC-Net, a dual-branch encoder–decoder architecture with double concatenation, designed for segmenting the liver and its lesions in CT images. Our network has a Y-shaped, similar to the Y-Net described in [28]. Nevertheless, the architecture proposed in [28] had been specifically designed to segment retinal layers and fluid pockets in ocular optical coherence tomography (OCT) images. This is accomplished by incorporating a second encoder branch that extracts spectral domain features in addition to the spatial encoder used in previous works. In our proposed network, two encoders can independently analyze input data and extract pertinent features, which are then propagated to a single decoder, which generates a 3D multi-class segmentation map, without the need to convert 2D medical images in 3D models.

Our Y-Net architecture leverages the feature aggregation capability provided by skip connections. Specifically, DEDC-Net utilizes encoder-to-decoder skip connections to integrate low- and high-level features meticulously. Our results demonstrate that the double concatenation via these skip connections enhances interconnectivity between final feature maps at each resolution level. This approach enables efficient feature reuse and addresses class imbalance more effectively than traditional architectures like 3DUNet and 3DVNet, as well as more recent ones like UNETR and SwinTr. Indeed, the LiTS dataset is affected by high-class imbalance, indicating that the tumor class we are attempting to identify and label appears substantially less frequently than others.

Table 3
Ablation analysis of our method for improving tumor segmentation on the LiTS dataset (15 test volumes). AG: Attention Gate, ED: Encoder-Decoder architecture with skip connections, EE: Encoder-Encoder architecture with double skip connections.

Encoder 1	Encoder 2	AG	ED	EE	Class	ASSD	VS	SN	SP	PR	DS
VGGNet	VGGNet	–	√	–	Liver	0.558 ± 0.318	0.944 ± 0.036	0.933 ± 0.075	0.996 ± 0.001	0.856 ± 0.052	0.889 ± 0.037
VGGNet	VGGNet	–	–	√		0.841 ± 16.867	0.921 ± 0.043	0.921 ± 0.093	0.995 ± 0.002	0.848 ± 0.056	0.878 ± 0.043
VGGNet	VGGNet	√	√	–		1.149 ± 1.82	0.889 ± 0.096	0.894 ± 0.168	0.995 ± 0.001	0.837 ± 0.063	0.851 ± 0.092
VGGNet	VGGNet	√	–	√		0.629 ± 0.343	0.932 ± 0.035	0.914 ± 0.106	0.995 ± 0.001	0.844 ± 0.048	0.873 ± 0.048
ResNet	ResNet	–	√	–		1.121 ± 1682	0.910 ± 0.063	0.901 ± 0.147	0.995 ± 0.001	0.853 ± 0.041	0.869 ± 0.092
ResNet	ResNet	–	–	√		0.771 ± 0.574	0.918 ± 0.031	0.923 ± 0.099	0.994 ± 0.001	0.833 ± 0.029	0.873 ± 0.054
ResNet	ResNet	√	√	–		0.866 ± 0.54	0.942 ± 0.031	0.948 ± 0.052	0.995 ± 0.002	0.856 ± 0.046	0.898 ± 0.033
ResNet	ResNet	√	–	√		0.948 ± 1.167	0.923 ± 0.062	0.91 ± 0.118	0.996 ± 0.002	0.872 ± 0.039	0.886 ± –0.071
VGGNet	ResNet (ours)	–	√	–		0.873 ± 0.915	0.918 ± 0.038	0.935 ± 0.085	0.993 ± 0.003	0.806 ± 0.063	**0.898 ± 0.031**
VGGNet	ResNet	–	–	√		1.082 ± 0.884	0.915 ± 0.053	0.935 ± 0.055	0.993 ± 0.003	0.789 ± 0.064	0.854 ± 0.044
VGGNet	ResNet	√	√	–		0.894 ± 0.953	0.922 ± 0.048	0.898 ± 0.122	0.994 ± 0.003	0.810 ± 0.065	0.847 ± 0.079
VGGNet	ResNet	√	–	√		0.996 ± 0.941	0.910 ± 0.038	0.888 ± 0.132	0.994 ± 0.004	0.817 ± 0.054	0.845 ± 0.072
VGGNet	VGGNet	–	√	–	Lesions	18.485 ± 30.814	0.612 ± 0.360	0.276 ± 0.309	0.999 ± 0.009	0.394 ± 0.356	0.289 ± 0.283
VGGNet	VGGNet	–	–	√		15.012 ± 14.271	0.485 ± 0.280	0.262 ± 0.297	0.999 ± 0.001	0.394 ± 0.356	0.240 ± 0.193
VGGNet	VGGNet	√	√	–		23.349 ± 30.037	0.431 ± 0.329	0.240 ± 0.293	0.999 ± 0.001	0.310 ± 0.320	0.202 ± 0.220
VGGNet	VGGNet	√	–	√		11.013 ± 13.497	0.438 ± 0.418	0.310 ± 0.329	0.999 ± 0.001	0.540 ± 0.375	0.313 ± 0.308
ResNet	ResNet	–	√	–		13.445 ± 24.796	0.594 ± 0.335	0.466 ± 0.364	0.999 ± 0.001	0.438 ± 0.332	0.413 ± 0.307
ResNet	ResNet	–	–	√		5.228 ± 14.845	0.534 ± 0.378	0.313 ± 0.342	0.999 ± 0.001	0.471 ± 0.295	0.320 ± 0.296
ResNet	ResNet	√	√	–		16.560 ± 21.176	0.591 ± 0.261	0.42 ± 0.352	0.998 ± 0.001	0.353 ± 0.294	0.336 ± 0.269
ResNet	ResNet	√	–	√		9.064 ± 8.803	0.641 ± 0.329	0.449 ± 0.345	0.998 ± 0.003	0.433 ± 0.343	0.401 ± 0.30
VGGNet	ResNet (ours)	–	√	–		12.168 ± 12.666	0.673 ± 0.282	0.570 ± 0.323	0.998 ± 0.003	0.464 ± 0.310	**0.461 ± 0.274**
VGGNet	ResNet	–	–	√		26.197 ± 34.108	0.647 ± 0.292	0.398 ± 0.355	0.998 ± 0.002	0.335 ± 0.300	0.328 ± 0.277
VGGNet	ResNet	√	√	–		12.681 ± 14.445	0.518 ± 0.395	0.372 ± 0.381	0.998 ± 0.002	0.363 ± 0.331	0.305 ± 0.299
VGGNet	ResNet	√	–	√		28.313 ± 35.704	0.502 ± 0.338	0.236 ± 0.284	0.999 ± 0.002	0.365 ± 0.323	0.235 ± 0.235

Table 4
Segmentation results of our proposed network on two unseen datasets, 3D-IRCARDb and COMET.

Dataset	Class	ASSD	VS	SN	SP	PR	DS
3D-IRCARDb	Liver	0.648 ± 0.958	0.910 ± 0.058	0.956 ± 0.104	0.989 ± 0.104	0.835 ± 0.083	0.886 ± 0.077
	Lesions	4.934 ± 8.806	0.730 ± 0.291	0.580 ± 0.300	0.999 ± 0.001	0.787 ± 0.202	0.629 ± 0.271
COMET	Liver	1.553 ± 3.478	0.900 ± 0.113	0.885 ± 0.198	0.995 ± 0.001	0.852 ± 0.031	0.854 ± 0.129
	Lesions	14.382 ± 30.499	0.582 ± 0.347	0.429 ± 0.298	0.998 ± 0.004	0.524 ± 0.366	0.435 ± 0.323

Due to their under-representation during training, uncommon classes could eventually end up being ignored. This is precisely the case for tumor segmentation, where tumor diameters are frequently considerably smaller than those of the liver.

Furthermore, the LiTS paper highlights an even greater challenge in accurately segmenting small tumors, amplifying the difficulties faced in such tasks [49]. The LiTS dataset, which includes cases with tumors measuring only a few voxels in diameter, poses a substantial obstacle for segmentation efforts. This challenge is exacerbated by the small number of surrounding pixels available to define tumor borders, particularly problematic in low-resolution images of 256 × 256 pixels in axial slices, as encountered in our research. To empirically evaluate this issue, we evaluated our proposed networks, on a selected subset of the OSLO-COMET dataset that included cases with exceptionally small tumors (diameters ranging from 3 mm to 10 mm). The results from the OSLO-COMET dataset, specifically for small-sized tumors, yielded significantly lower performance metrics, as evidenced by the DS of 0.461. These outcome underscore the inherent difficulties in detecting and accurately segmenting smaller lesions, attributable to challenges in visibility, substantial variation in shape, appearance, location, and the performance metrics' heightened sensitivity to minor inaccuracies for small-scale targets. Furthermore, medical imaging's inherent susceptibility to noise and artifacts complicates the identification of these small tumors. Despite these hurdles, DEDC-Net has demonstrated superior performance in liver segmentation tasks. The inclusion of residual blocks within our architecture's multi-encoder framework plays a pivotal role in extracting detailed features at each encoder layer, enhancing our model's efficacy. To demonstrate our network's generalization capability in clinical settings, we evaluated the 3DYNet-ED across multiple datasets, including LiTS, 3D-IRCADb, and OSLO-COMET. Our findings, which show the model's superior performance in lesion segmentation on the IRCADB-01 dataset with a Dice Score of 0.629, affirm the model's robustness and adaptability across different liver and tumor medical datasets.

The Attention Gate (AG) mechanism [21], integrated into encoder–decoder architectures, enhances key features passed through skip connections and is widely recognized for its utility in various medical image segmentation tasks [21,23,50]. However, its performance is closely tied to the training data distribution, highlighting its data-driven nature. In the context of the LiTS dataset, incorporating AGs into traditional architectures like 3DUNet and 3DVNet did not result in significant improvements in liver and tumor segmentation, aligning with findings from previous studies [51]. In our ablation analysis, we observed that integrating AGs into our Encoder-Encoder variant networks significantly improved segmentation scores, although it did not achieve the highest scores for liver and tumor segmentation. AGs filter and prioritize feature activations for the decoder, complementing the residual blocks in our multi-encoder configuration to extract detailed features at each encoder layer. This synergy between residual connections and AGs has shown promising results, warranting further investigation.

3D multi-class segmentation of the liver and its tumors is a challenging task, not only because of the imbalance in medical datasets but also due to the computational complexity required to preserve 3D spatial information [52]. Although our network showed promising outcomes for liver segmentation, the lesion segmentation results were relatively low. This may be due to GPU memory limitations necessitating the down-sampling of volumes to a size of 256 × 256 × 64 for further processing. This resampling reduces the spatial resolution and detail available for accurate lesion identification, leading to lower segmentation performance. This explains the variation in our results compared to those in the literature, but still allows for a fair comparison with existing methods. Future research should rigorously evaluate the impact of various AG mechanisms on our architecture to better handle class imbalances in liver and tumor segmentation. Enhancing this aspect is crucial for improving network reliability in imbalanced class scenarios. Additionally, we plan to investigate the performance of networks with two encoders based on different architectures and extend testing to additional medical image segmentation datasets.

7. Conclusions

In this paper, we introduce DEDC-Net, a dual-branch encoder–decoder architecture where we use both long- and short-range skip connections. The model is specifically designed for segmenting liver CT volumes, enhancing feature reusability. Through both qualitative and quantitative evaluations, our approaches outperform existing state-of-the-art liver segmentation methods on the LiTS dataset. An ablation study examining various encoder backbones (VGG19 and ResNet) and the integration of attention mechanisms in the decoder stage revealed that *DEDC-Net* delivers superior liver segmentation accuracy, achieving mean DS of 0.898 liver segmentation without the need for additional attention gates before the up-sampling operation of each decoder block. Our research also tackles the challenge of lesion segmentation, which is crucial for effective treatment planning and diagnosis, alongside liver segmentation. Despite the complexities of lesion segmentation our network reached the highest DS of 0.629 on the unseen dataset IDCARDb-01. Future work will focus on thoroughly evaluating the impact of different attention mechanisms to improve our architecture's performance in managing class imbalances in liver and tumor segmentation. This is vital for enhancing network reliability in imbalanced scenarios. Moreover, exploring the efficiency of various network backbones with minimal hyper-parameters and extending tests to more medical imaging datasets could broaden our segmentation networks' applicability and performance in medical imaging.

CRediT authorship contribution statement

Gabriella d'Albenzio: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Yuliia Kamkova:** Writing – review & editing, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Rabia Naseem:** Writing – review & editing, Writing – original draft, Investigation, Conceptualization. **Mohib Ullah:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **Stefania Colonnese:** Writing – review & editing, Supervision, Project administration, Conceptualization. **Faouzi Alaya Cheikh:** Writing – review & editing, Supervision, Project administration, Investigation, Conceptualization. **Rahul Prasanna Kumar:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was financially supported by the Research Council of Norway through the ALive project [grant number 311393]; and the Department of Research and Development, Division of Emergencies and Critical Care, Oslo University Hospital, Rikshospitalet.

We would like to thank all who contributed to this project's success, with special appreciation to Dr. Kiran Raja for insightful comments on the manuscript, Eirik Gromholt Homlong for his guidance on implementing the networks, Dr. Egidijus Pelanis for providing the liver and tumors segmentations of the COMET data, Prof. Ole Jakob Elle and MD Prof. Bjørn Edwin for their valuable support and guidance on the clinical validation of our study. We are also grateful to Assoc. Prof. Rafael Palomar for his constant mentorship, encouragement, and instrumental support in bringing this paper to fruition.

References

[1] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, in: CA: A Cancer Journal for Clinicians, Wiley Online Library, 2021, pp. 209–249.

[2] T. Blum, H. Feußner, N. Navab, Modeling and segmentation of surgical workflow from laparoscopic video, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 400–407.

[3] B. Preim, C.P. Botha, Visual Computing for Medicine: Theory, Algorithms, and Applications, Newnes, 2013.

[4] P.F. Christ, M.E.A. Elshaer, F. Ettlinger, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, M. Armbruster, F. Hofmann, M. D'Anastasi, W.H. Sommer, S.-A. Ahmadi, B.H. Menze, Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016, Springer International Publishing, Cham, 2016, pp. 415–423.

[5] F. Lu, F. Wu, P. Hu, Z. Peng, D. Kong, Automatic 3D liver location and segmentation via convolutional neural network and graph cut, Int. J. Comput. Assist. Radiol. Surg. 12 (2) (2017) 171–182.

[6] J. Yamanaka, S. Saito, J. Fujimoto, Impact of preoperative planning using virtual segmental volumetry on liver resection for hepatocellular carcinoma, World J. Surg. 31 (2007) 1251–1257.

[7] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Springer International Publishing, 2015, pp. 234–241.

[8] S. Gul, M.S. Khan, A. Bibi, A. Khandakar, M.A. Ayari, M.E. Chowdhury, Deep learning techniques for liver and liver tumor segmentation: A review, Comput. Biol. Med. 147 (2022) 105620.

[9] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, D. Shen, Deep convolutional neural networks for multi-modality isointense infant brain image segmentation, NeuroImage 108 (2015) 214–224.

[10] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, H. Larochelle, Brain tumor segmentation with deep neural networks, Med. Image Anal. 35 (2017) 18–31.

[11] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.

[12] P. Hu, F. Wu, J. Peng, P. Liang, D. Kong, Automatic 3D liver segmentation based on deep learning and globally optimized surface evolution, Phys. Med. Biol. 61 (24) (2016) 8676.

[13] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-net: learning dense volumetric segmentation from sparse annotation, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016, Springer International Publishing, 2016, pp. 424–432.

[14] F. Milletari, N. Navab, S.-A. Ahmadi, V-Net: Fully convolutional neural networks for volumetric medical image segmentation, in: 2016 Fourth International Conference on 3D Vision, 3DV, IEEE, 2016, pp. 565–571.

[15] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, P.-A. Heng, H-DenseUNet: Hybrid densely connected unet for liver and tumor segmentation from CT volumes, IEEE Trans. Med. Imaging 37 (12) (2018) 2663–2674.

[16] W. Xu, H. Liu, X. Wang, Y. Qian, Liver segmentation in CT based on ResUNet with 3D probabilistic and geometric post process, in: 2019 IEEE 4th International Conference on Signal and Image Processing, ICSIP, IEEE, 2019, pp. 685–689.

[17] Z. Zhou, M.M.R. Siddiquee, N. Tajbakhsh, J. Liang, UNet++: Redesigning skip connections to exploit multiscale features in image segmentation, IEEE Trans. Med. Imaging 39 (6) (2019) 1856–1867.

[18] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, J. Wu, UNet 3+: A full-scale connected unet for medical image segmentation, in: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2020, pp. 1055–1059.

[19] L. Qian, X. Zhou, Y. Li, Z. Hu, UNet#: A unet-like redesigning skip connections for medical image segmentation, 2022, arXiv:2205.11759.

[20] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, 2016, arXiv:1409.0473.

[21] O. Oktay, J. Schlemper, L.L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N.Y. Hammerla, B. Kainz, B. Glocker, D. Rueckert, Attention U-net: Learning where to look for the pancreas, 2018, arXiv:1804.03999.

[22] N. Abraham, N.M. Khan, A novel focal tversky loss function with improved attention U-net for lesion segmentation, in: 2019 IEEE 16th International Symposium on Biomedical Imaging, ISBI 2019, 2019, pp. 683–687.

[23] J. Zhang, Z. Jiang, J. Dong, Y. Hou, B. Liu, Attention gate resu-net for automatic MRI brain tumor segmentation, IEEE Access 8 (2020) 58533–58545.

[24] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, D. Rueckert, Attention gated networks: Learning to leverage salient regions in medical images, Med. Image Anal. 53 (2019) 197–207.

[25] H. Jiang, T. Shi, Z. Bai, L. Huang, AHCNet: An application of attention mechanism and hybrid connection for liver tumor segmentation in CT volumes, IEEE Access 7 (2019) 24898–24909.

[26] Q. Jin, Z. Meng, C. Sun, H. Cui, R. Su, RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans, Front. Bioeng. Biotechnol. 8 (2020) 1471.

[27] L. Jiang, J. Ou, R. Liu, Y. Zou, T. Xie, H. Xiao, T. Bai, Rmau-net: Residual multi-scale attention u-net for liver and tumor segmentation in ct images, Comput. Biol. Med. 158 (2023) 106838.

[28] A. Farshad, Y. Yeganeh, P. Gehlbach, N. Navab, Y-net: A spatiospectral dual-encoder network for medical image segmentation, in: Medical Image Computing and Computer Assisted Intervention – MICCAI 2022, Springer Nature Switzerland, 2022, pp. 582–592.

[29] K.-N. Wang, S.-X. Li, Z. Bu, F.-X. Zhao, G.-Q. Zhou, S.-J. Zhou, Y. Chen, SBCNet: Scale and boundary context attention dual-branch network for liver tumor segmentation, IEEE J. Biomed. Health Inf. 28 (5) (2024) 2854–2865, http://dx.doi.org/10.1109/JBHI.2024.3370864.

[30] H. Liu, J. Yang, C. Jiang, S. He, Y. Fu, S. Zhang, X. Hu, J. Fang, W. Ji, S2DA-net: Spatial and spectral-learning double-branch aggregation network for liver tumor segmentation in CT images, Comput. Biol. Med. 174 (2024) 108400.

[31] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2015, arXiv:1409.1556.

[32] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, IEEE, 2016, pp. 770–778.

[33] A. Mohammed, S. Yildirim, I. Farup, M. Pedersen, Ø. Hovde, Y-net: A deep convolutional neural network for polyp detection, 2018, arXiv:1806.01907.

[34] S. Mehta, E. Mercan, J. Bartlett, D. Weaver, J.G. Elmore, L. Shapiro, Y-net: joint segmentation and classification for diagnosis of breast biopsy images, in: Medical Image Computing and Computer Assisted Intervention – MICCAI 2018, Springer International Publishing, 2018, pp. 893–901.

[35] G. Klambauer, T. Unterthiner, A. Mayr, S. Hochreiter, Self-normalizing neural networks, Adv. Neural Inf. Process. Syst. 30 (2017).

[36] J. Ma, J. Chen, M. Ng, R. Huang, Y. Li, C. Li, X. Yang, A.L. Martel, Loss odyssey in medical image segmentation, Med. Image Anal. 71 (2021) 102035.

[37] S.A. Taghanaki, Y. Zheng, S.K. Zhou, B. Georgescu, P. Sharma, D. Xu, D. Comaniciu, G. Hamarneh, Combo loss: Handling input and output imbalance in multi-organ segmentation, Comput. Med. Imaging Graph. 75 (2019) 24–33.

[38] Å.A. Fretland, A.M. Kazaryan, B.A. Bjørnbeth, K. Flatmark, M.H. Andersen, T.I. Tønnessen, G.M.W. Bjørnelv, M.W. Fagerland, R. Kristiansen, K. Øyri, et al., Open versus laparoscopic liver resection for colorectal liver metastases (the Oslo-CoMet study): study protocol for a randomized controlled trial, Trials 16 (2015) 1–10.

[39] P. Bilic, P. Christ, H.B. Li, E. Vorontsov, A. Ben-Cohen, G. Kaissis, A. Szeskin, C. Jacobs, G.E.H. Mamani, G. Chartrand, et al., The liver tumor segmentation benchmark (LiTS), Med. Image Anal. 84 (2023) 102680.

[40] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, PyTorch: An imperative style, high-performance deep learning library, in: Advances in Neural Information Processing Systems, Vol. 32, Curran Associates, Inc., 2019, pp. 8026–8037.

[41] K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification, in: 2015 IEEE International Conference on Computer Vision, ICCV, IEEE Computer Society, 2015, pp. 1026–1034.

[42] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2017, arXiv:1412.6980.

[43] I. Loshchilov, F. Hutter, SGDR: Stochastic gradient descent with warm restarts, 2017, arXiv:1608.03983.

[44] A. Abdollahi, B. Pradhan, A. Alamri, VNet: An end-to-end fully convolutional neural network for road extraction from high-resolution remote sensing data, IEEE Access 8 (2020) 179424–179436.

[45] A. Turečková, T. Tureček, Z. Komínková Oplatková, A. Rodríguez-Sánchez, Improving CT image tumor segmentation through deep supervision and attentional gates, Front. Robot. AI 7 (2020) 106.

[46] X. Han, MR-based synthetic CT generation using a deep convolutional neural network method, Med. Phys. 44 (4) (2017) 1408–1419, http://dx.doi.org/10.1002/mp.12155.

[47] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, M. Wang, Swin-unet: Unet-like pure transformer for medical image segmentation, in: European Conference on Computer Vision, Springer, 2022, pp. 205–218.

[48] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H.R. Roth, D. Xu, Unetr: Transformers for 3d medical image segmentation, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 574–584.

[49] P. Bilic, P.F. Christ, E. Vorontsov, G. Chlebus, H. Chen, Q. Dou, C.-W. Fu, X. Han, P.-A. Heng, J. Hesser, et al., The liver tumor segmentation benchmark (lits), 2019, arXiv preprint arXiv:1901.04056.

[50] V.-T. Pham, T.-T. Tran, P.-C. Wang, P.-Y. Chen, M.-T. Lo, EAR-UNet: A deep learning-based approach for segmentation of tympanic membranes from otoscopic images, Artif. Intell. Med. 115 (2021) 102065.

[51] M. Chung, J. Lee, S. Park, C.E. Lee, J. Lee, Y.-G. Shin, Liver segmentation in abdominal CT images via auto-context neural network and self-supervised contour attention, Artif. Intell. Med. 113 (2021) 102023.

[52] D.T. Kushnure, S.N. Talbar, HFRU-net: High-level feature fusion and recalibration unet for automatic liver and tumor segmentation in CT images, Comput. Methods Programs Biomed. 213 (2022) 106501.