Computerized Medical Imaging and Graphics 116 (2024) 102414

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

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

Accurate segmentation of liver tumor from multi-modality non-contrast images using a dual-stream multi-level fusion framework

Chenchu Xu [a,b], Xue Wu [a], Boyan Wang [c,*], Jie Chen [a], Zhifan Gao [d], Xiujian Liu [d,**], Heye Zhang [d]

[a] *Artificial Intelligence Institute, Anhui University, Hefei, China*
[b] *Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei, China*
[c] *Department of Computer Science and Technology, Tsinghua University, Beijing, China*
[d] *School of Biomedical Engineering, Sun Yat-sen University, Shenzhen, China*

ARTICLE INFO

Keywords:
Non-contrast liver tumor segmentation
Multi-modality images fusion
Dual-stream network

ABSTRACT

The use of multi-modality non-contrast images (i.e., T1FS, T2FS and DWI) for segmenting liver tumors provides a solution by eliminating the use of contrast agents and is crucial for clinical diagnosis. However, this remains a challenging task to discover the most useful information to fuse multi-modality images for accurate segmentation due to inter-modal interference. In this paper, we propose a dual-stream multi-level fusion framework (DM-FF) to, for the first time, accurately segment liver tumors from non-contrast multi-modality images directly. Our DM-FF first designs an attention-based encoder–decoder to effectively extract multi-level feature maps corresponding to a specified representation of each modality. Then, DM-FF creates two types of fusion modules, in which a module fuses learned features to obtain a shared representation across multi-modality images to exploit commonalities and improve the performance, and a module fuses the decision evidence of segment to discover differences between modalities to prevent interference caused by modality's conflict. By integrating these three components, DM-FF enables multi-modality non-contrast images to cooperate with each other and enables an accurate segmentation. Evaluation on 250 patients including different types of tumors from two MRI scanners, DM-FF achieves a Dice of 81.20%, and improves performance (Dice by at least 11%) when comparing the eight state-of-the-art segmentation architectures. The results indicate that our DM-FF significantly promotes the development and deployment of non-contrast liver tumor technology.

1. Introduction

Multi-modality non-enhanced image-based liver tumor segmentation is a time-efficient, safe, and cost-effective solution for clinical diagnosis and treatment. This approach entails the use of different types of MRIs (i.e., T1FS, T2FS, and DWI) without the injection of contrast agents, which nonetheless offer sufficient information regarding tumor characteristics that are comparable to contrast-enhanced images (Vu et al., 2018; Zhao et al., 2020). For instance, T1FS enables the observation of bleeding in liver tumors (Vu et al., 2018), T2FS is often useful in diagnosing benign lesions (Vu et al., 2018), and DWI displays details of hepatocellular carcinoma and cholangiocarcinoma (Zhang et al., 2023). Compared with clinical contrast-enhanced image-based liver tumor segmentation, the application of multi-modality non-enhanced images affords the following advantages: firstly, it avoids fatal reactions resulting from contrast agents (Schieda et al., 2018), particularly the potential toxicity observed in patients with impaired renal function; secondly, it reduces extra imaging time resulting from contrast agent injection and waiting (Gao et al., 2023; Xu et al., 2020); and lastly, it circumvents the cost of contrast agents (Xu et al., 2023) (see Fig. 1).

Currently, two multi-modality methods (Zhao et al., 2020, 2021) that introduce contrast-enhanced images as intermediate are attempting to segment liver tumors without contrast agents. However, there is still a lack of investigation into multi-modality methods based solely on non-contrast images. In the absence of guidance provided by the contrast-enhanced images, the integration of different non-enhanced modalities of liver tumors becomes critical to non-enhanced image-based methods. Such methods should explore the shared representation and identify common features, while minimizing differences among modalities to improve segmentation performance. At the same time, these methods should retain specific information regarding each modality during the learning process and maintain the integrity of the specified semantics of each modality to avoid interference between

* Corresponding author.
** Co-corresponding author.
E-mail addresses: wby000000@mail.tsinghua.edu.cn (B. Wang), liuxj86@mail.sysu.edu.cn (X. Liu).

https://doi.org/10.1016/j.compmedimag.2024.102414
Received 29 January 2024; Received in revised form 10 May 2024; Accepted 28 June 2024
Available online 3 July 2024

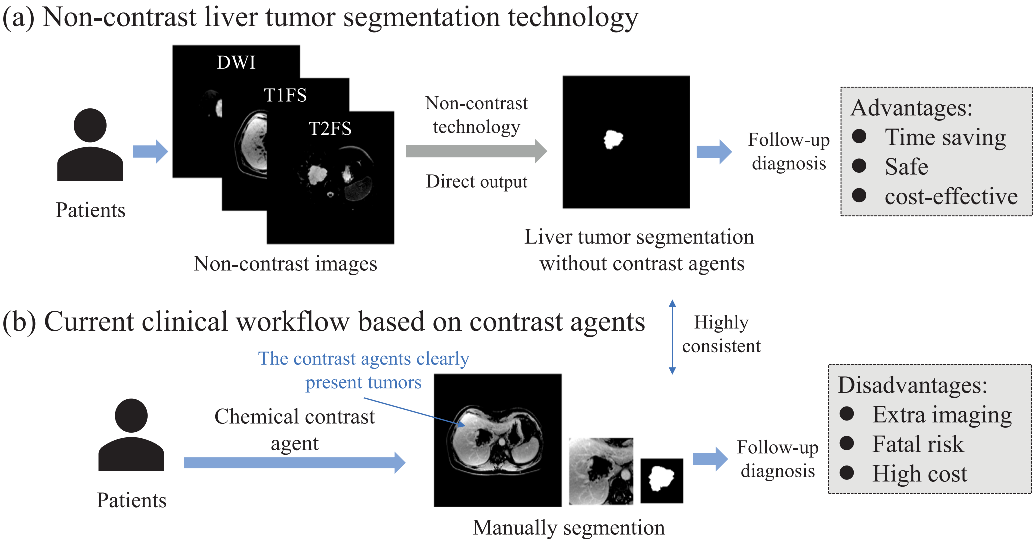


Fig. 1. A potential clinical alternative is proposed to eliminate the use of contrast agents in enhanced MRI. The method is a time-saving, safe and cost-effective solution that simplifies the clinical workflow and can provide detailed information on liver tumors.

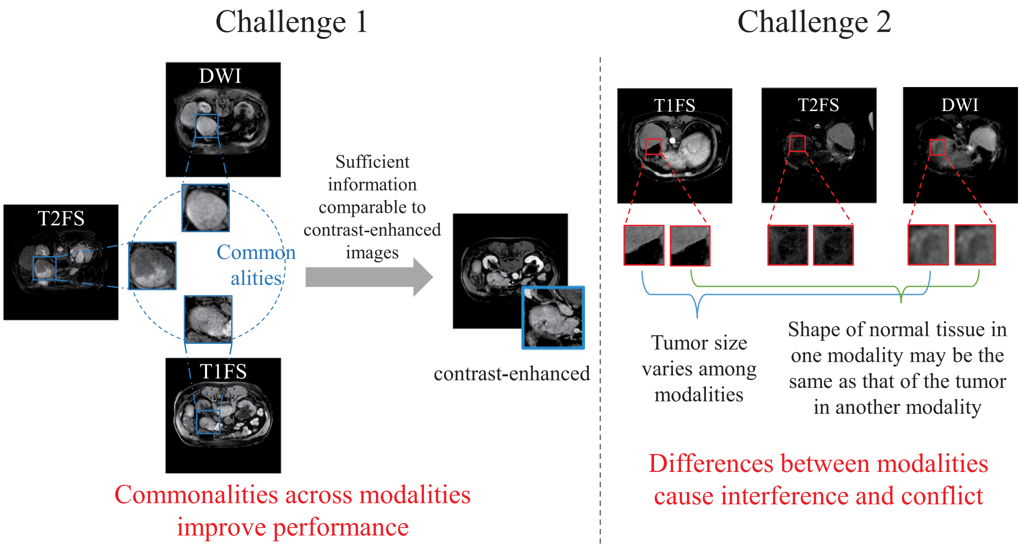

Challenge 1
DWI
T2FS
T1FS
Common alities
Sufficient information comparable to contrast-enhanced images
contrast-enhanced
Commonalities across modalities improve performance
Challenge 2
T1FS
T2FS
DWI
Tumor size varies among modalities
Shape of normal tissue in one modality may be the same as that of the tumor in another modality
Differences between modalities cause interference and conflict

Fig. 2. Multi-modality non-contrast images contain sufficient information about tumor characteristics comparable to contrast-enhanced images. But without the guidance of contrast-enhanced images, how integrating different modalities of liver tumors without contrast agents has become critical.

modalities. For instance, as illustrated in Fig. 2, the appearance of the same tumor in different modalities may vary, and the shape of normal tissue in one modality might be similar to that of the tumor in another modality.

This study presents a novel dual-stream multi-level fusion framework (DM-FF) to accurately segment liver tumors from multi-modality non-contrast images directly. Our approach explores commonalities and identifies conflicts across and between multi-modality non-contrast images, thus enhancing performance and preventing interference. Specifically, our DM-FF embeds a receptive field attention (RFA) module into an attention-based encoder–decoder architecture that effectively extracts multi-level feature maps representing a specified representation of each modality. Furthermore, a layer fusion module expands the multi-axis vision transformer to construct a shared representation with long-term dependency across multi-modality non-contrast images, obtaining commonality information. Additionally, a decision fusion module employs evidence theory (Shafer, 1976) to handle uncertainties and conflicts between images, resulting in a reliable segmentation outcome.

In summary, our work offers the following key contributions:

- A new contrast-free segmentation method is proposed to exploit the commonalities and differences between multi-modality images to enable directly segmenting liver tumors from non-contrast images. This eliminates the use of contrast agents along with their related risks, extra time, and cost.
- We propose a Receptive Field Attention Module (RFA) to efficiently extract fine-grained feature mappings with high-level semantic information and low-level detail information, while reducing the number of channels and efficiently decreasing the computational complexity.
- The newly designed multi-modality image fusion method that combines the two fusion strategies to learn the shared representations between modalities by Layer Fusion Module (LFM) while

simultaneously learning the specific semantics of each modality by Decision Fusion Module(DFM), thereby better fusing the modalities and improving the segmentation accuracy.

2. Related work

2.1. Existing deep-learning-based on liver tumor methods

The current focus of liver tumor research is to utilize multi-modality and non-contrast techniques to provide more accurate and safer diagnostic tools. However, they rely to some extent on contrast-enhanced images. For example, Almotairi et al. (2020) used a semantic pixel-wise classification for road scenes is adopted and modified to fit liver CT segmentation and classification. Hänsch et al. (2022) used an anisotropic 3D U-Net architecture and a multi-model training strategy to segment liver tumors from contrast-enhanced MRI images. Jiang et al. (2023) proposed a novel Residual Multi-scale Attention U-Net for liver and tumor segmentation in CT. In addition, these studies introduce contrast-enhanced images as intermediate methods to attempt to segment liver tumors without contrast. Xiao et al. (2019) combined enhanced radiological features with a discriminator that segmented liver tumor from non-contrast T2WI using a generator. Xu et al. (2021) used deep reinforcement learning to synthesize contrast-enhanced MRI using non-contrast T2WI MRI. Zhao et al. (2021) first predicts liver tumors by non-contrast multi-modality MRI detection, and then improves the prediction accuracy by incorporating enhanced MRI into the discriminator. Xu et al. (2022) proposed a new ternary knowledge transferred teacher-student DRL that uses teacher network to learn contrast-enhanced tumors to guide student network to learn non-contrast tumors for contrast-free liver tumor detection. There is still a lack of methods to completely eliminate the need for enhanced images and rely only on non-contrast images to achieve multi-modality liver tumor segmentation. Therefore, we propose a model that fully and exclusively utilizes the information of multi-modality non-contrast images to achieve accurate segmentation of liver tumors.

2.2. Medical multi-modality segmentation methods

Multi-modality is increasingly used in medical imaging, where it provides multiple pieces of information about the target (i.e., tumor, organ or tissue) (Zhou et al., 2019). Currently, the segmentation methods for medical multi-modality images are mainly in the following two ways. On the one hand, it is to take all modalities as a whole at the input and improve the segmentation network to achieve accurate segmentation. For example, Ibtehaz and Rahman (2020) proposed MultiResUNet that modifies the classical U-Net to enable excellent performance in segmenting multi-modality medical images. Wang et al. (2021) proposed TransBTS that combines the advantages of 3D CNN and Transformer to both model local context and learn global information associations. On the other hand, it is to create inter-modal relationships at the feature level. For example, Xing et al. (2022) proposed NestedFormer based on a transformer-based multi-encoder and single-decoder architecture that performs nested multi-modality fusion of high-level representations of different modalities. Zhang et al. (2022) proposed mmFormer to generate modality-invariant features by performing local and global context modeling within each modality while establishing long-range dependencies across modalities. Wang et al. (2023) proposed MFCNet to fuse CT and PET image information by a multi-modality fusion down-sampling block with residual structure. Although these methods can establish relationships between modalities, it ignores information specific to each modality. Moreover, due to the lack of guidance provided by contrast-enhanced images, this leads to the accumulation of different modal errors in the extraction path. To address this limitation, we propose a segmentation model using a two-stream multilevel fusion approach. The model allows us to establish relationships between each modality while maintaining the integrity of each modality-specific semantics and avoiding inter-modal interference.

2.3. Semantic segmentation methods

In the era of deep learning, computer vision researchers are dedicated to improving the performance of semantic segmentation and related tasks. Despite existing works, such as Oktay et al. (2018) and Fu et al. (2019), which integrate self-attention mechanisms into segmentation models to enhance efficiency and accuracy by capturing global contextual information, they can result in high computational costs and cannot effectively utilize local contextual information. Additionally, semantic segmentation networks proposed in Zhao et al. (2017) and Chen et al. (2018) combine multi-scale feature extraction, multi-pathways, multi-scale pathway fusion, and expansion convolution to capture rich contextual information in an image and improve segmentation results. However, these methods still cannot fully meet the requirements of extracting tumor information from non-contrast images because, compared to enhanced images, features in non-contrast images are often not very clear. To better capture and utilize image feature information, we propose a method that combines multi-scale feature extraction, dilated convolution, and attention mechanisms. This method aims to extract rich feature information from non-contrast images, and reduce the high computational cost problem of self-attention by using multi-axis attention mechanisms.

3. Methodology

The DM-FF is a liver tumor segmentation model that utilizes multi-modality non-contrast images, including T1FS, T2FS, and DWI as inputs, and fuses them together to produce segmentation results. The DM-FF comprises three main modules: (1) a Receptive Field Attention module that employs four parallel convolution channels with globally refined attention to achieve a specified representation of each non-contrast modality; (2) a Layer Fusion module that incorporates bottom-up splicing with the transformer to fuse modalities at the feature level and obtain a shared representation across modalities for segmentation; and (3) a Decision Fusion module that measures the uncertainty of the pixel class to fuse modalities at the decision level and mitigate interference between modalities. Integration of the three modules enables the exploitation of commonalities and differences across and between modalities for accurate segmentation, as illustrated in Fig. 3. It is important to note that the ground truth used in this study was obtained by radiologists from contrast-enhanced images.

3.1. Overall architecture

The architecture of our proposed model is illustrated in Fig. 3. Independent encoder for each modality process three layers of features extracted from the backbone network through corresponding Receptive Field Attention module. These feature maps are employed in decision and layer fusion processes. In the decision fusion process, the features obtained from the encoder are fed to the corresponding decoder to generate the initial segmentation results for each modality. These results are subsequently fused in the Decision Fusion module to produce corresponding segmentation results. The independent encoder–decoder for each modality prevents the interference of different modality's feature extraction paths, thus avoiding the accumulation of errors. Higher-level representation enhances the robustness and reliability of multi-modality fusion. In the layer fusion process, the feature maps of each modality are combined at each layer from the encoder and sent to the corresponding decoder. The initial fusion and segmentation results arising from this process are sent to the top Layer Fusion module. In the Layer Fusion module, the features of the three modalities at this layer, the previous fusion result and the segmentation result, are fused. These fused and segmented results are then fed to the next layer. Finally, the feature fusion result and the segmentation result are obtained at the bottom level. This top-down gradual fusion of modalities can achieve modality fusion at the feature level, which is conducive to the

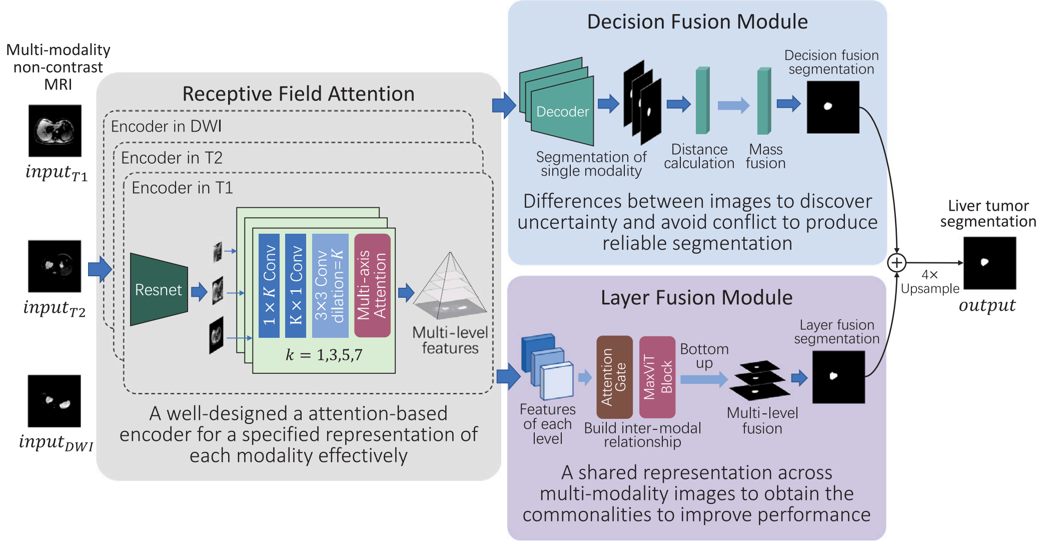

Multi-modality non-contrast MRI
$input_{T1}$
$input_{T2}$
$input_{DWI}$
Receptive Field Attention
Encoder in DWI
Encoder in T2
Encoder in T1
Resnet
1 × K Conv
K × 1 Conv
3×3 Conv dilation=K
Multi-axis Attention
k = 1,3,5,7
Multi-level features
A well-designed a attention-based encoder for a specified representation of each modality effectively
Decision Fusion Module
Decoder
Segmentation of single modality
Distance calculation
Mass fusion
Decision fusion segmentation
Differences between images to discover uncertainty and avoid conflict to produce reliable segmentation
Liver tumor segmentation
4× Upsample
output
Layer Fusion Module
Features of each level
Attention Gate
MaxViT Block
Build inter-modal relationship
Bottom up
Multi-level fusion
Layer fusion segmentation
A shared representation across multi-modality images to obtain the commonalities to improve performance

Fig. 3. The DM-FF includes three novel modules. (1) Receptive field attention module achieves a specified representation of each non-contrast modality respectively. (2) Layer fusion module fuse modalities at the feature level and obtains a shared representation across modalities for segmentation. (3) Decision fusion module fuse modalities at the decision level and avoid interference between modalities. By integrating three modules, DM-FF enables exploiting commonalities and differences across and between modalities for accurate segmentation.

interaction of information between modalities and the acquisition of correlation between modalities. The resulting fusion is highly reliable and accurate. The final segmentation result is obtained by summing up the two segmentation results.

We combine two losses as the basic loss $\mathcal{L}_{Basic}$ which is the mean of binary cross entropy (BCE) loss and Intersection over Union(IOU) loss, that is $\mathcal{L}_{Basic} = \mathcal{L}_{BCE} + L_{IoU}$. The loss function for model training is:

$$\mathcal{L} = \alpha\mathcal{L}_{Basic}\left(out_l, \hat{y}\right) + \beta\mathcal{L}_{Basic}\left(out_d, \hat{y}\right) + \gamma\mathcal{L}_{Basic}\left(y, \hat{y}\right) \tag{1}$$

where $\hat{y}$ represents ground truth, y represents final segmentation result, out_l and out_d represents the segmentation results of the layer fusion module and the decision fusion module, respectively. α, β and γ are the parameters that control the proportion.

3.2. Receptive field attention module

The Receptive Field Attention module is a crucial component that enhances global refinement across multiple scales and levels of each modality. The module is structured as a multi-branch system that comprises convolution layers with convolution kernels in various sizes, and the Multi-Axis Attention module (Tu et al., 2022) is integrated into each branch. This configuration enables the extraction of fine-grained feature maps with high-level semantic information and low-level detail information while simultaneously reducing the number of channels and effectively mitigates computational complexity.

As shown in Fig. 3, the feature maps of the three different layers obtained by each modality through their respective backbone networks are forwarded to the corresponding Receptive Field Attention modules for additional global refinement of each scale and connect the outputs. The specific composition of the Receptive Field Attention module, shown in Fig. 4, consists of four branches that use the regular convolution at different scales, the 3 × 3 dilated convolution with different dilated rate, and the multi-axis attention module. Meanwhile, the regular convolutional layers are replaced with the corresponding $1 \times n$ and $n \times 1$ convolutions to reduce the parameters and deeper nonlinear layers. By using dilated convolution kernels with different dilation rates on each branch, the perceptual field expands without loss of resolution, allowing the model to obtain multi-scale contextual information with less computational effort. The above process can be expressed using the following formula:

$$x_k = \begin{cases} \mathrm{Conv}_{1\times1}(x), & if\ \ k = 1 \\ \mathrm{MAA}(\mathrm{DConv}_k(\mathrm{Conv}_k(x))), & if\ \ k = 3, 5, 7 \end{cases} \tag{2}$$

where $k \in [3, 5, 7]$, Conv_k denotes the $1 \times k$ and $k \times 1$ convolution. DConv_k denotes the 3×3 convolution with the dilated rate of k.

After applying regular and dilated convolution in each branch, we incorporate a multi-axis attention module to extract global dependencies and local representations. The Multi-Axis Attention module decomposes full-scale attention into two sparse forms, local and global, and divides the input into non-overlapping local blocks firstly to perform self-attention in local space, thus achieving local interaction and recovering the original form after computation. Next, the input is aggregated into different local blocks based on a global expansion, and self-attention is performed in the local space corresponding to the global. This step allows for global interaction, and subsequently reconstructs the original shape after computation. Therefore, the Multi-axis Attention module enables global-local space interactions at arbitrary input resolutions while maintaining linear complexity.

Finally, the global refinement results of the four branches are merged and sent to successive convolution layers to obtain the module's output, expressed as formula:

$$x_{out} = \mathrm{Conv}\left(\mathrm{Concat}\left(x_1, x_3, x_5, x_7\right)\right) \tag{3}$$

The output is used for the subsequent decision fusion of the decoder module and the layer fusion module.

3.3. Layer fusion module

The layer fusion module is a critical component that facilitates mining and constructing complex relationships between various modalities at the feature level. This module consists of bottom-up splicing, Spatial Attention, and Multi-Axis Vision Transformer (MaxViT) (Tu et al., 2022). It gradually integrates modal features by linking the features from various layers of each modality in the backbone network. This approach facilitates integrating and comprehensively using multi-modality images at the feature level.

Layer fusion begins with an initial fusion of the features of all layers in all modalities by the MaxViT decoder, and then fuses the features of each layer in a bottom-up manner. The fusion of each layer

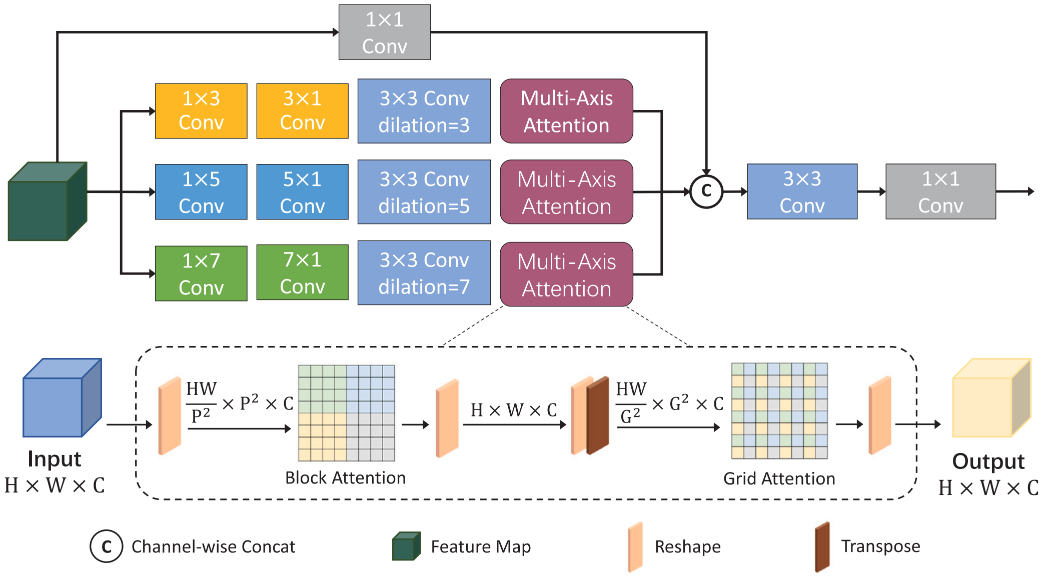


Fig. 4. The Receptive Field Attention Module achieves a specific representation of each non-contrast modality, respectively. The Receptive Field Attention module is composed of three parts: (1) Extracting branches through convolution at different scales for enhanced global feature refinement; (2) Utilizing Dilated Convolution to expand the perceived field without sacrificing resolution; (3) Implementing Multi-Axis Attention module to extract global dependencies and local representations. In this image example, the values of P and G are 4, whereas in actual use they are 7.

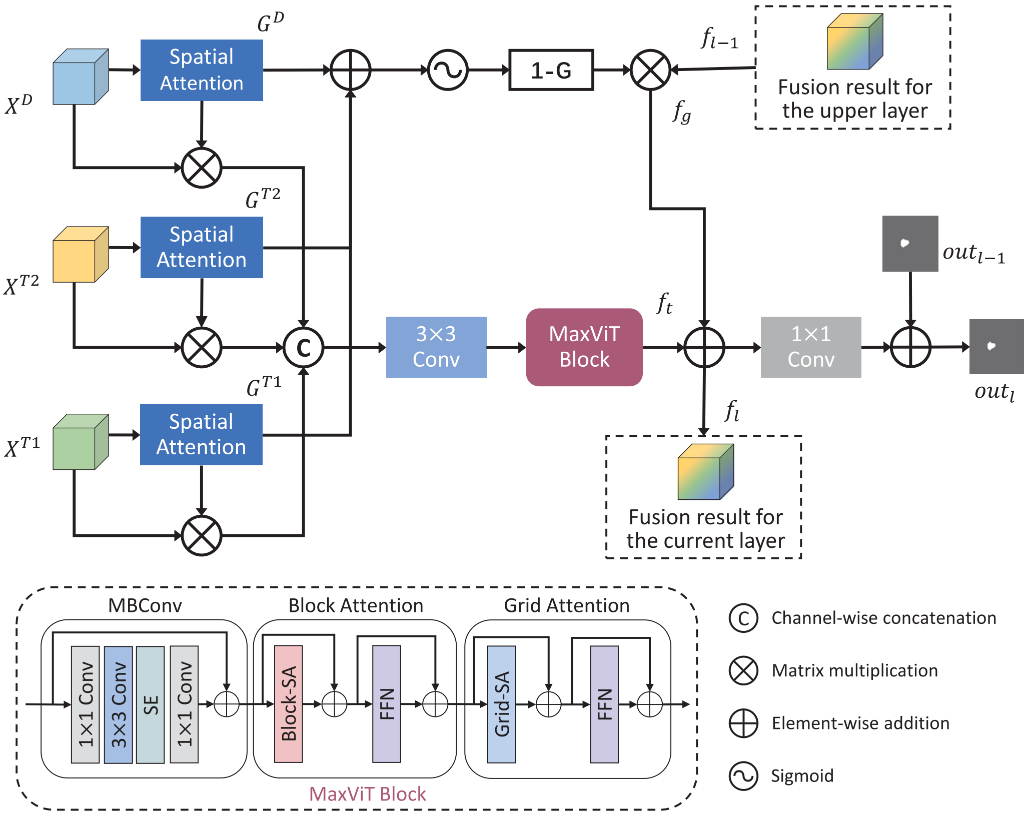


Fig. 5. The Layer Fusion module fuses modalities at the feature level and acquires the shared representation across modalities for segmentation. The Layer Fusion Module is composed of three parts: (1) Using spatial attention as a gate to spotlight the feature information of the modalities at this layer and diminish irrelevant information in the higher layer; (2) Multi-Axis Vision Transformer (MaxViT) for establishing inter-modal relationships for feature fusion; (3) Fusing the feature information of both layers to yield the fused information and segmentation results at this layer.

is shown in Fig. 5, where the features of each mode of the layer, which are obtained by the encoder, are used as input. The feature information of each modality is denoted as X^{T1}, X^{T2}, X^D and the useful information $G \in \{G^{T1}, G^{T2}, G^D\}$ is separately highlighted by spatial attention. Then it is multiplied by its corresponding modality separately to preserve the useful information of each modality. Meanwhile, the information f_{l-1} from the upper-layer modal fusion is filtered backward by $1 - G$, retaining the information f^g that is not redundant between the upper-layer and the current layer. The calculation of f^g as follow:

$$f^g = f_{l-1}\left(1 - \text{softmax}\left(G^{T1} + G^{T2} + G^D\right)\right) \tag{4}$$

To obtain the fusion result F^t of this layer through concatenating the useful information of each modality and establishing the long-term correlation among modalities using MaxViT. The calculation of f^t as follow:

$$f^t = \text{MaxViT}\left(\text{Conv}\left(\text{Concat}\left(X^{T1}G^{T1}, X^{T2}G^{T2}, X^D G^D\right)\right)\right) \tag{5}$$

The fusion result f^t is added to f^g to obtain the final fusion result f_l for the layer. After applying 1 × 1 convolution, it is added to the fusion segmentation result from the upper layer to obtain the fusion segmentation result out_l for the layer.

Layer Fusion module provides fusion of dual aspectual information for multi-modality images. The first is the feature information of each modality in the current layer, and after highlighting the important feature information, the information is merged to determine the relationship of each modality in that layer. The second is the fused information of each modality in the upper layer. The propagation of useful information in the upper layer is controlled by the use of gates, and the propagation of useless information can be effectively suppressed. This ensures adequate fusion of modal information at the feature level and reduces noise in the fusion process.

3.4. Decision fusion module

The decision fusion module is responsible for establishing correlations between different modalities at the decision level and resolving uncertainties and conflicts when combining information from each modality. The module comprises uncertainty quantification and fusion techniques, and an evidence fusion strategy is employed to obtain independently extracted path information, achieve uncertainty quantification for each modality's pixel class by mass function. Finally, Dempster's rule (Huang et al., 2021) is applied to achieve fusion of modalities at the decision level. This approach reduces error accumulation and enhances the reliability and robustness of the multi-modality fusion.

The initial segmentation results in f^{t1}, f^{t2}, and f^d are obtained at each modality using a separate encoder–decoder. These results are then combined to obtain the initial fusion feature f_a, which is subsequently fed into the decision fusion layer. In this layer, a mass is initially assigned to each K-class and the entire class set Ω based on the distance between the feature vector of each pixel and the center of the I-prototype. For a given pixel x, each prototype p_i is considered as evidence whose reliability decreases with the Euclidean distance d_i between x and p_i. Each prototype p_i is considered to have a membership degree u_{ik} for each class ω_k with the constraint $\sum_{k=1}^{K} u_{ik} = 1$. The mass function induced by the prototype p_i is:

$$m_i(\{\omega_k\}) = a_i u_{ik} exp(-\gamma_i d_i^2), k = 1, \ldots, K \tag{6}$$

$$m_i(\Omega) = 1 - a_i exp(-\gamma_i d_i^2), \tag{7}$$

The mass functions induced by the I-prototype are then combined by means of the Dempster's rule:

$$m = \bigoplus_{i=1}^{I} m_i \tag{8}$$

A belief function quantifying the uncertainty of each pixel category is obtained, and the final fusion results out_{DF} is obtained according to the result of this function.

Algorithm 1: Training DM-FF

```
Input: Multi-modality non-contrast liver images
       $\{\mathcal{X}^{T1}, \mathcal{X}^{T2}, \mathcal{X}^{D}\}$; Segmentation label y
Output: Tumor segmentation results of input
Initialize framework
for Each epoch do
    for Each step do
        begin forward propagation:
        Stage1: Feature extraction
        for modality $i$ in $\{T1, T2, D\}$ do
            $\{r_1^i, r_2^i, r_3^i\} = \text{Resnet}(\mathcal{X}^i)$
            for feature $r$ in $\{r_1, r_2, r_3\}$ do
                $\mathcal{R} = \text{RFA}(r)$
            end
        end
        Stage2: Decision fusion of segmentation
        for modality $i$ in $\{T1, T2, D\}$ do
            $f^i = \text{Decoder}(\mathcal{R}_1^i, \mathcal{R}_2^i, \mathcal{R}_3^i)$
        end
        $out_D = \text{DFM}(f^{T1}, f^{T2}, f^D)$
        Stage3: Layer fusion of segmentation
        $out_0, f_0 = \text{Decoder}(\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3)$
        for $l$ in L do
            $out_l, f_l = \text{LFM}(\mathcal{R}_l^{T1}, \mathcal{R}_l^{T2}, \mathcal{R}_l^{D}, out_{l-1}, f_{l-1})$
        end
        Stage4: Get final segmentation result
        $\hat{y} = out_D + out_L$
        begin backward propagation:
        Update framework according to Eq. (1)
    end
end
```

4. Experiments

4.1. Dataset

Liver tumor dataset: The dataset consists of 250 liver tumor cases, including 150 cases of Hemangioma and 100 cases of Hepatocellular Carcinoma. All participants underwent the initial standard clinical protocol for liver magnetic resonance imaging examination. The non-contrast liver MRI were acquired using a 3-T MRI system (GE Signa) with the image size of 256 × 256, the slice thickness of 6 mm, and the pixel pitch of 1.2882 × 1.2882 mm. The segmentation ground truth is derived from tumor contours delineated on the enhanced images by an experienced radiologist with seven years of experience in liver MRI. The images were enhanced using the Fat Saturation sequence after the intravenous injection (20 s to 10 min, triphase) of a gadolinium-based contrast agent and the specific gadolinium-based contrast agent used was 0.1 mmol/kg and was obtained from Bayer Schering Pharma AG in Berlin, Germany. The size of the obtained liver enhanced MRI images is 512 × 512, the section thickness is 4 mm, and the pixel spacing is 0.6441 × 0.6441 mm. To maintain consistency with the non-enhanced images, the enhanced images were re-sliced and resized. According to the ground truth, the average length of the major axis of the lesions was 3.2 cm (Standard Deviation: 2.9 cm; Minimum value: 0.8 cm; Maximum: 22.0 cm), and the tumor area on average occupied 1.6% of the image area.

LLD-MMRI2023 dataset (Lou et al., 2024): The dataset contains 498 annotated multi-phase liver lesions from the same number of patients, with the lesions classified into seven categories. Each lesion has eight phases. For the purposes of this study, 100 hepatocellular carcinomas and 45 hemangiomas were selected for analysis. Four phases were employed for each lesion: non-contrast, T2-weighted imaging, diffusion-weighted imaging, and venous. The venous phase was utilized to obtain the segmentation ground truth, which was then used for the study. Furthermore, due to the differing volumes of each phase, the three non-contrast phases of the same case were resampled to the same voxel intervals as the venous phase. Finally, the slices with the segmentation ground truth values were retained.

4.2. Implementation details

We provide a detailed overview of the model's structure in Section 3. Additionally, the convolutional layer within the backbone network maintains a consistent channel count of 32. The backbone network used is a Res2Net (Gao et al., 2019) with a setting of $26w \times 4s$. We perform the five-fold cross-validation test to train our DM-FF model for performance evaluation and comparison. It divides the dataset into five equal parts, using four parts as training sets each time and one part as a test set. This process is repeated five times, with different test sets each time. Finally, the evaluation results of each time are averaged to obtain the final evaluation results of the model. The images are resized to 224 × 224 for training, and resized back to their original size at the end of testing. The data enhancement technique used is similar to that described in the work by Kim et al. (2021). We utilized the Adam optimizer (Kingma and Ba, 2014) with an initial learning rate of 10^{-4}, which was fine-tuned from a range of options: [1e−3, 1e−4, 1e−5]. To facilitate the learning rate decay, we employed a polynomial function with a decay (Chen et al., 2017) factor of $(1-(\frac{iter}{iter_{m}ax})^{0.9})$. For the batch size, we opted for 16 after considering several values: [8, 16, 32, 64]. The model implementation was done using Pytorch (Paszke et al., 2019), while the model training was performed on a GeForce RTX 3090 GPU.

4.3. Evaluation metrics

The DM-FF using five well-recognized metrics, including Dice Similarity Coefficient (Dice), Intersection Over Union (IOU), Mean Absolute Error (MAE), Precision, and Recall. These metrics are defined as follows:

Dice similarity coefficient (Dice) evaluates the accuracy of image segmentation results by comparing the similarity between the real results and the predicted results. Dice is defined as:

$$Dice = \frac{1}{N}\sum_{i=1}^{N}\frac{2\left|P_i \cap G_i\right|}{|P_i| + |G_i|} \tag{9}$$

where N is the number of samples, G_i is the true output value of sample i, and P_i is the predicted output value. When the Dice coefficient is close to 1, it means that the accuracy of image segmentation is high.

Intersection over Union (IoU) measures the degree of overlap in a segmentation by calculating the ratio of the intersection to the union between the predicted segmentation and the ground truth segmentation. IoU is defined as:

$$IoU = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|P_i \cap G_i\right|}{\left|P_i \cup G_i\right|} \tag{10}$$

where N, i, G_i and P_i have the same indication as above. IoU values range from 0 to 1, with 1 indicating exact agreement between predicted and ground truth tumor region.

Mean Absolute Error (MAE) measures the accuracy of the predicted segmentation region by calculating the mean absolute error between the prediction result and ground-truth. MAE is defined as:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|P_i - G_i\right| \tag{11}$$

where N, i, G_i and P_i have the same indication as above. If the MAE is smaller, it means that the image segmentation result is more accurate.

Precision measures the proportion of correctly predicted tumor regions in the whole predicted region. Precision is defined as:

$$Precision = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FP_i} \tag{12}$$

where TP represents the area correctly predicted as tumor and FP represents the area incorrectly predicted as tumor. The Precision values range from 0 to 1, with higher values indicating more accurately predicted areas by the segmentation model.

Recall measures the percentage of correct predictions across the entire tumor area in ground truth. Recall is defined as:

$$Recall = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FN_i} \tag{13}$$

where TP represents the area correctly predicted as tumor and FN represents the area incorrectly predicted as background. The Recall values range from 0 to 1, with higher values indicating that the segmentation model captures the tumor area better.

4.4. Experimental results and analysis

Comprehensive experiments have demonstrated that DM-FF is capable of generating accurate segmentation in multi-modality non-contrast images. DM-FF achieves a Dice of 81.20%, an IOU of 71.02%, a Precision of 82.32%, a Recall of 84.08%, and an MAE of 0.17% in the Liver tumor dataset. Similarly, DM-FF achieves a Dice of 80.58%, an IOU of 69.44%, a Precision of 77.39%, a Recall of 88.23%, and an MAE of 0.53% in the LLD-MMRI2023 dataset. These result indicates that DM-FF has great potential to promote the non-contrast liver tumor segmentation technology applied in clinical setting and eliminate the health risks associated with contrast-agents injections.

4.4.1. Segmentation result

The experimental results from Fig. 6, Tables 1 and 2 indicate that our DM-FF is able to accurately segment liver tumor from non-contrast multi-modality magnetic resonance imaging. The visualization in Fig. 6 shows that our results are in high consistent with the ground truth obtained by the radiologist from the contrast-enhanced images. Although tumors are locating in different regions and presenting different sizes and shapes, our method can accurately predict their boundaries by combining multi-modality commonalities and differences. For example, in Case 1 and Case 3, the tumor size is small and the tumor shape in both T1FS and DWI modalities is extremely different from the correct shape. DM-FF accurately segments these two cases. In addition, Tables 1 and 2 show the high performance of our DM-FF on various segmentation metrics. In the Liver tumor dataset, DM-FF achieved 81.20%, 71.02%, 82.32%, 84.08% and 0.17% of Dice, IOU, Precision, Recall and MAE, respectively. Furthermore, DM-FF achieved 80.58%, 69.44%, 77.39%, 88.23% and 0.53% of Dice, IOU, Precision, Recall and MAE, respectively, in the LLD-MMRI2023 dataset. This confirms its effectiveness in contrast-free liver segmentation.

4.4.2. Performance comparison

As Fig. 6, Tables 1 and 2 illustrate, both experimental results indicate our method out-performs these state-of-the-art methods. We compared our DM-FF with eight state-of-the-art methods, which can be further grouped into two categories: (1) Five standard object segmentation methods: PSPnet (Zhao et al., 2017), DeeplabV3+ (Chen et al., 2018), UNet++ (Zhou et al., 2018), ResUNet (Zhang et al., 2018), and Attention-UNet (Oktay et al., 2018). (2) Three recent non-contrast liver segmentation methods are Weakly-Supervised Teacher-Student network (WSTS Zhang et al., 2021 2021), United Adversarial Learning (UAL Zhao et al., 2021 2021), and Multi-scale Intermediate Multi-modal Fusion Network (MIMFNet Pan et al., 2021 2021). All of these methods were trained using the default network structure and parameters listed in reference. In addition, these methods can only segment

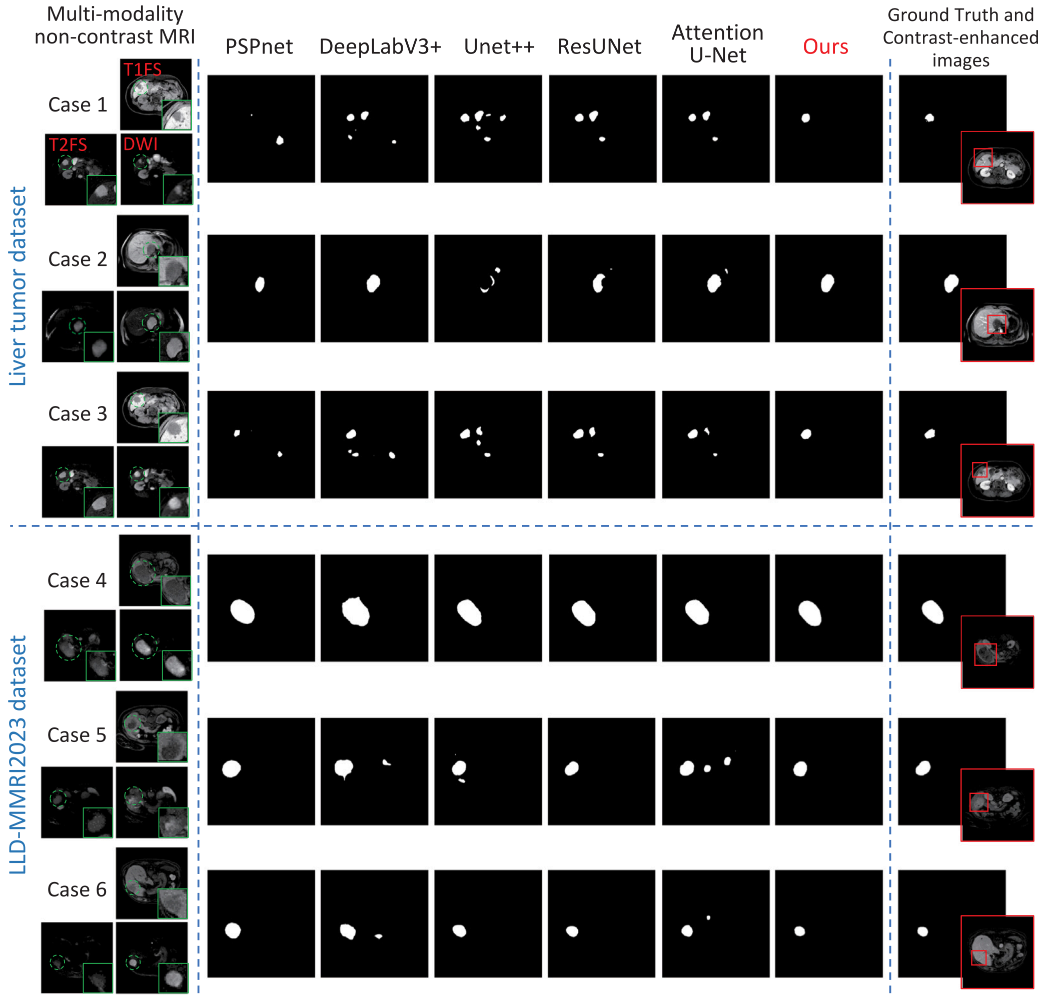


Fig. 6. Our DM-FF accurately segments liver tumors from non-contrast images, and outperforms all five comparative methods.

Table 1
Performance comparison on Liver tumor dataset. The results show that our proposed DM-FF is superior to eight state-of-the-art methods for liver tumor segmentation.

Method	Dice ↑	IoU ↑	Precision ↑	Recall ↑	MAE ↓
PSPnet	48.30 ± 29.61	36.69 ± 26.12	52.45 ± 34.53	53.33 ± 32.88	0.59 ± 0.43
DeepLabV3+	65.13 ± 26.15	53.31 ± 25.70	67.21 ± 29.71	72.68 ± 28.43	0.40 ± 0.53
UNet++	70.92 ± 24.58	60.24 ± 27.47	68.72 ± 27.57	84.10 ± 23.75	0.40 ± 0.45
ResUNet	63.64 ± 28.86	52.99 ± 29.81	59.84 ± 31.96	82.05 ± 27.10	0.50 ± 0.55
Attention-UNet	62.06 ± 30.61	51.70 ± 30.23	61.21 ± 34.74	75.14 ± 31.24	0.51 ± 0.64
WSTS	–	43.15 ± 19.49	59.71 ± 24.00	60.88 ± 23.05	–
MIMFNetde	–	49.76 ± 12.82	68.27 ± 14.48	64.46 ± 13.11	–
UAL	–	49.69 ± 18.80	65.39 ± 16.69	67.42 ± 15.78	–
Our	**81.20** ± **16.28**	**71.02** ± **18.36**	**82.32** ± **15.91**	**84.08** ± **19.59**	**0.17** ± **0.18**

Table 2
Performance comparison on LLD-MMRI2023 dataset. The results show that our proposed DM-FF is superior to five state-of-the-art methods for liver tumor segmentation.

Method	Dice ↑	IoU ↑	Precision ↑	Recall ↑	MAE ↓
PSPnet	54.41 ± 26.59	41.79 ± 24.18	45.78 ± 26.11	76.25 ± 30.26	1.70 ± 0.97
DeepLabV3+	51.72 ± 25.48	38.62 ± 21.69	41.49 ± 23.10	77.44 ± 31.72	1.75 ± 0.73
UNet++	62.87 ± 23.73	50.01 ± 23.97	53.20 ± 25.09	86.73 ± 24.53	1.53 ± 1.38
ResUNet	74.23 ± 14.95	61.20 ± 17.42	66.62 ± 20.21	**90.68** ± **13.39**	1.04 ± 1.12
Attention-UNet	71.33 ± 17.83	58.29 ± 19.92	63.89 ± 20.78	88.56 ± 19.22	1.03 ± 0.98
Our	**80.58** ± **13.49**	**69.44** ± **16.05**	**77.39** ± **17.01**	88.23 ± 14.71	**0.53** ± **0.34**

liver tumors from non-contrast liver MRIs for tumor segmentation without the use of contrast agents.

The quantitative analysis of the segmentation results is presented in Tables 1 and 2. These results show that DM-FF outperforms five comparative standard segmentation methods for evaluation of all five metrics in two datasets, and outperforms three comparative liver segmentation methods for evaluation of three metrics in the Liver tumor dataset. In the Liver tumor dataset, our DM-FF improves the Dice by

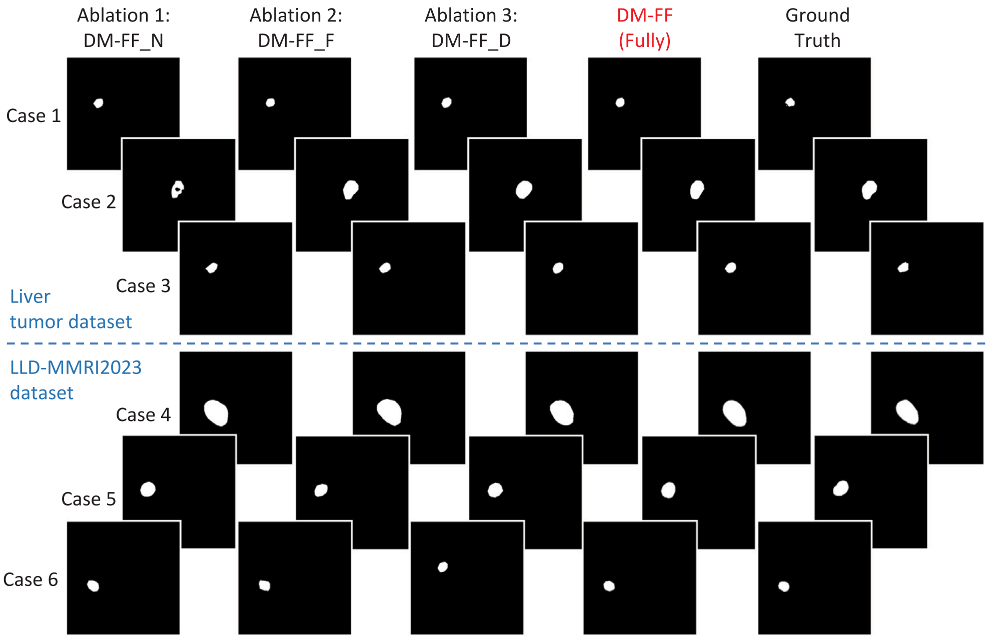


Fig. 7. Segmentation results of ablation study.

Table 3
The quantitative evaluation of ablation study. These results show that each module of the newly designed DM-FF plays a role in liver tumors segmentation.

Dataset	Method	Dice ↑	IoU ↑	Precision ↑	Recall ↑	MAE ↓
Liver tumor	DM-FF_N	73.66 ± 19.07	63.16 ± 21.25	73.12 ± 18.59	77.82 ± 22.44	0.25 ± 0.21
	DM-FF_F	77.86 ± 21.21	67.88 ± 23.06	78.17 ± 20.20	83.67 ± 22.72	0.21 ± 0.23
	DM-FF_D	72.21 ± 21.65	60.25 ± 21.72	67.64 ± 22.59	**84.91 ± 26.30**	0.29 ± 0.29
	DM-FF	**81.20 ± 16.28**	**71.02 ± 18.36**	**82.32 ± 15.91**	84.08 ± 19.59	**0.17 ± 0.18**
LLD-MMRI2023	DM-FF_N	77.20 ± 14.74	65.00 ± 16.76	73.95 ± 17.46	86.05 ± 18.04	0.67 ± 0.49
	DM-FF_F	67.51 ± 20.49	54.25 ± 20.86	73.23 ± 20.09	73.16 ± 28.42	0.90 ± 0.66
	DM-FF_D	76.78 ± 18.61	65.43 ± 19.97	**77.47 ± 21.63**	81.65 ± 20.08	0.59 ± 0.49
	DM-FF	**80.58 ± 13.49**	**69.44 ± 16.05**	77.39 ± 17.01	**88.23 ± 14.71**	**0.53 ± 0.34**

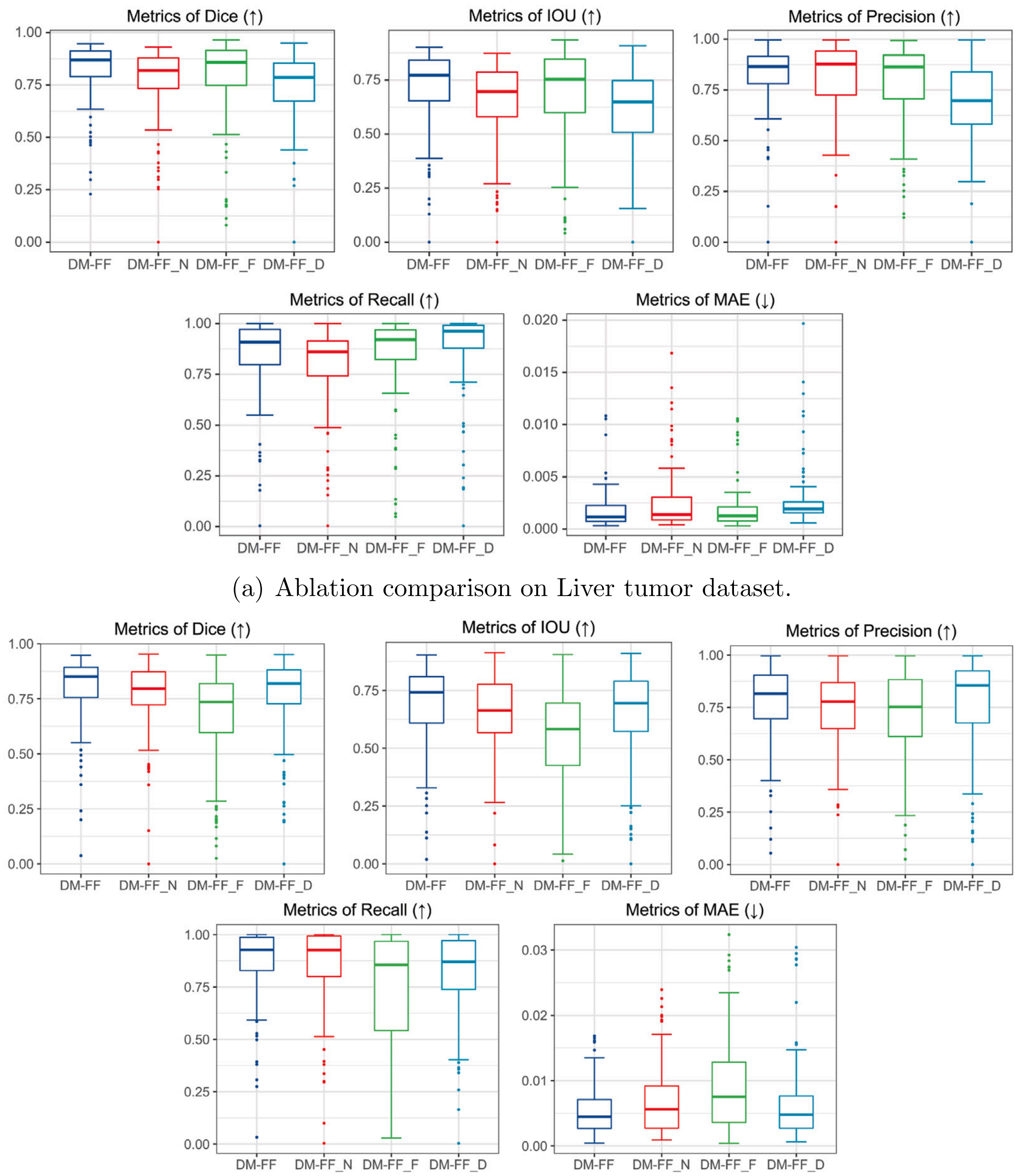
11%–33%, the IOU by 11%–35%, the Precision by 14%–30%, and the MAE by 0.23–0.42%. And in the LLD-MMRI2023 dataset, our DM-FF improves the Dice by 6%–28%, the IOU by 8%–30%, the Precision by 10%–35%, and the MAE by 0.5–1.22%. These results demonstrate that the segmentation of DM-FF are more accurate and have higher consistency with annotations than other methods. More intuitively, Fig. 6 visually shows that compared with other segmentation methods. From Fig. 6, we can see that DM-FF improves the accuracy of image segmentation and avoids mis-segmentation on background regions that similar to the object. Specifically, in Cases 1 and 3, due to the tumor morphology is similar to the surrounding normal tissues in the individual modalities, this misleads other methods into segmenting the normal tissues while segmenting the tumor, and our method focus on the tumor. All the outperformance is because our DM-FF combines two types of fusion strategies to learn the specified semantics of each modality, but also learn a shared representation between modalities, thus improving the fusion accuracy.

4.4.3. Ablation study

Advantage of Receptive Field Attention module. Fig. 7 and Table 3 of our study demonstrate that the Receptive Field Attention module has a significant impact on improving the results of liver tumor segmentation. To evaluate the impact of the Receptive Field Attention module, we created model version DM-FF_N, which is identical to DM-FF but does not include this module. Our results demonstrate that DM-FF yields better tumor segmentation compared to DM-FF_N. Specifically, in Liver tumor and LLD-MMRI2023 datasets, DM-FF improves the Dice by 7.54% and 3.38%, respectively, compared to the model without the Receptive Field Attention module. The improved performance of DM-FF can be attributed to the ability of the Receptive Field Attention module to refine global features in each modality, at different scales and levels, and extract more useful information. This demonstrates the importance of extracting modality features during the multi-modality non-contrast liver tumor segmentation process.

Advantage of Decision Fusion module. Fig. 7 and Table 3 of our study demonstrate that the Decision Fusion module has a more significant improvement on the segmentation results of liver tumors. To evaluate the impact of the Decision Fusion module, we designed a version of our model, called DM-FF_F, which only uses the Layer Fusion module for feature fusion and missing the decoder part associated with Decision Fusion. Our results demonstrate that DM-FF yields better tumor segmentation compared to DM-FF_F. Specifically, in Liver tumor and LLD-MMRI2023 datasets, DM-FF improves the Dice by 3.34% and 13.07%, respectively, compared to DM-FF_F which only uses the feature fusion. This is due to the Layer Fusion module is primarily focused on establishing inter-modality relationships at the feature level, without considering the specific characteristics of each modality. This may result in incomplete or erroneous feature representations, which can compromise the final segmentation performance.

Advantage of Layer Fusion module. Fig. 7 and Table 3 of our study demonstrate that the Layer Fusion module has substantially enhanced the segmentation outcomes of liver tumors. To examine the contribution of the Decision Fusion module, we designed a variant of our model called DM-FF_D, which retains the decision fusion approach but removes the decoder part associated with the Layer Fusion module. Our results demonstrate that DM-FF yields better tumor segmentation



(a) Ablation comparison on Liver tumor dataset.

(b) Ablation comparison on LLD-MMRI2023 dataset.

Fig. 8. The comparison results of the ablation study indicate that the Receptive Field Attention module, Decision Fusion module, and Layer Fusion module can enhance the segmentation performance. The best performance can be achieved by combining these three modules simultaneously.

compared to DM-FF_D. Specifically, in Liver tumor and LLD-MMRI2023 datasets, DM-FF improves the Dice by 8.99% and 3.80%, respectively, compared to DM-FF_D only using decision fusion. This is due to the Decision Fusion module may overlook the intricate relationships between different modalities and fail to capture the particular semantics of each modality, resulting in a significant loss of valuable feature information for interaction. In contrast, the Layer Fusion module establishes inter-modality relationships at the feature level while preserving the specific characteristics of each modality, leading to more effective feature fusion and improved segmentation outcomes.

5. Discussion and conclusion

Segmentation of liver tumors using multi-modality non-contrast magnetic resonance imaging is a time-efficient, secure, and cost-effective approach for clinical diagnosis and treatment, which is crucial for clinical research. The use of contrast agents presents several limitations: (1) Contrast agents require a long imaging time and may cause discomfort to patients. (2) The cost of contrast agents is high. (3) Patients with impaired renal function may experience potential toxicity, making it unsuitable for individuals with chronic kidney disease. In particular, gadolinium contrast retention has been associated with a 10%–15% incidence of contrast-induced nephropathy (Stacul et al., 2011), and over 17% of patients with liver disease have kidney disease (Lauenstein et al., 2015). Currently, there are studies that introduce contrast-enhanced images as intermediate are attempting to segment liver tumors without contrast agents. However, there is still a lack of investigation on multi-modality methods based on only non-contrast images.

Therefore, we propose a novel dual-stream multi-level fusion framework (DM-FF) that allows for the segmentation of liver tumors directly from non-contrast MRI images, completely eliminating the use of contrast agents. DM-FF has the following advantages: firstly, it constructs the Receptive Field Attention module to enhance the modal feature information at multiple scales and levels to improve the discriminability of tumor regions. Secondly, it constructs the Layer Fusion module to facilitate the exploration and construction of complex relationships between modalities through layer-wise fusion at the feature level. Finally, it constructs the Decision Fusion module to quantify the uncertainty of the segmentation results from modalities, establishing the correlation between modalities at the decision level and resolving issues of conflict.

To evaluate the performance of DM-FF, we conducted experiments on a dataset consisting of 250 cases and compared it with eight state-of-the-art methods using five well-established metrics. DM-FF outperformed the others, improving the Dice metric by 11%–33%. These results demonstrate that DM-FF provides accurate tumor segmentation results comparable to manual detection by radiologists from contrast-enhanced images. Additionally, the main modules of DM-FF were analyzed in the experiment to quantify their contributions to the segmentation accuracy, as shown in Table 3 and Fig. 8. The research results demonstrate that DM-FF achieves precise segmentation of liver tumors by integrating these three modules and enabling collaboration between multimodality non-contrast images.

Despite successfully promoting progress in the field of non-contrast liver segmentation, our study still has several limitations. Firstly, it relies on data from a single medical center, which may limit the generalization of research results to other centers with different imaging protocols and patient populations. To address this, future work may involve implementing a federated learning approach that combines multiple center datasets. By sharing and aggregating data across multiple centers, more robust and generalizable models can be developed. Furthermore, the training process relies heavily on obtaining ground truth tumor segmentation from a large number of enhanced images. Obtaining these enhanced images may not always be easily accessible in clinical practice. To overcome this, future research can explore semi-supervised learning methods that utilize labeled and unlabeled data to learn from a larger dataset and reduce reliance on enhanced images.

CRediT authorship contribution statement

Chenchu Xu: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Data curation, Conceptualization. **Xue Wu:** Writing – original draft, Visualization, Validation, Methodology, Investigation, Data curation, Conceptualization. **Boyan Wang:** Visualization, Validation, Methodology, Investigation, Data curation, Conceptualization. **Jie Chen:** Validation, Investigation, Data curation, Conceptualization. **Zhifan Gao:** Writing – review & editing, Investigation, Data curation, Conceptualization. **Xiujian Liu:** Investigation, Data curation. **Heye Zhang:** Investigation.

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.

Data availability

The authors do not have permission to share data.

Acknowledgments

This work was supported in part by The National Natural Science Foundation of China (62106001, U1908211), The University Synergy Innovation Program of Anhui Province, China (GXXT-2021-007), The Anhui Provincial Natural Science Foundation, China (2208085Y19), The POS RADIOAMICA project funded by the Italian Minister of Health (CUP: H53C22000650006), and The Anhui Province Clinical Medical Research Transformation Special Project, China (202304295107 020031).

References

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.

Chen, L.-C., Papandreou, G., Schroff, F., Adam, H., 2017. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision. ECCV, pp. 801–818.

Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., Lu, H., 2019. Dual attention network for scene segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3146–3154.

Gao, S.-H., Cheng, M.-M., Zhao, K., Zhang, X.-Y., Yang, M.-H., Torr, P., 2019. Res2net: A new multi-scale backbone architecture. IEEE Trans. Pattern Anal. Mach. Intell. 43 (2), 652–662.

Gao, Z., Guo, Y., Zhang, J., Zeng, T., Yang, G., 2023. Hierarchical perception adversarial learning framework for compressed sensing MRI. IEEE Trans. Med. Imaging.

Hänsch, A., Chlebus, G., Meine, H., Thielke, F., Kock, F., Paulus, T., Abolmaali, N., Schenk, A., 2022. Improving automatic liver tumor segmentation in late-phase MRI using multi-model training and 3D convolutional neural networks. Sci. Rep. 12 (1), 12262.

Huang, L., Denœux, T., Tonnelet, D., Deca-zes, P., Ruan, S., 2021. Deep PET/CT fusion with Dempster-Shafer theory for lymphoma segmen-tation. In: Machine Learning in Medical Imaging: 12th International Workshop, MLMI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, Sep-tember 27, 2021, Proceedings 12. Springer, pp. 30–39.

Ibtehaz, N., Rahman, M.S., 2020. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. Neural Netw. 121, 74–87.

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

Kim, T., Lee, H., Kim, D., 2021. Uacanet: Uncertainty augmented context attention for polyp segmentation. In: Proceedings of the 29th ACM International Conference on Multimedia. pp. 2167–2175.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Lauenstein, T., Ramirez-Garrido, F., Kim, Y.H., Rha, S.E., Ricke, J., Phongkitkarun, S., Boettcher, J., Gupta, R.T., Korpraphong, P., Tanomkiat, W., et al., 2015. Nephrogenic systemic fibrosis risk after liver magnetic resonance imaging with gadoxetate disodium in patients with moderate to severe renal impairment: results of a prospective, open-label, multicenter study. Invest. Radiol. 50 (6), 416.

Lou, M., Ying, H., Liu, X., Zhou, H.-Y., Zhang, Y., Yu, Y., 2024. SDR-former: A siamese dual-resolution transformer for liver lesion classification using 3D multi-phase imaging. arXiv preprint arXiv:2402.17246.

Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al., 2018. Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999.

Pan, C., Zhou, P., Tan, J., Sun, B., Guan, R., Wang, Z., Luo, Y., Lu, J., 2021. Liver tumor detection via a multi-scale intermediate multi-modal fusion network on MRI images. In: 2021 IEEE International Conference on Image Processing. ICIP, IEEE, pp. 299–303.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al., 2019. Pytorch: An imperative style, high-performance deep learning library. Adv. Neural Inf. Process. Syst. 32.

Schieda, N., Blaichman, J.I., Costa, A.F., Glikstein, R., Hurrell, C., James, M., Jabehdar Maralani, P., Shabana, W., Tang, A., Tsampalieros, A., et al., 2018. Gadolinium-based contrast agents in kidney disease: a comprehensive review and clinical practice guideline issued by the Canadian Association of Radiologists. Can. J. Kidney Health Dis. 5, 2054358118778573.

Shafer, G., 1976. A Mathematical Theory of Evidence, vol. 42, Princeton University Press.

Stacul, F., van der Molen, A.J., Reimer, P., Webb, J.A., Thomsen, H.S., Morcos, S.K., Almén, T., Aspelin, P., Bellin, M.-F., Clement, O., et al., 2011. Contrast induced nephropathy: updated ESUR contrast media safety committee guidelines. Eur. Radiol. 21, 2527–2541.

Tu, Z., Talebi, H., Zhang, H., Yang, F., Milanfar, P., Bovik, A., Li, Y., 2022. Maxvit: Multi-axis vision transformer. In: Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXIV. Springer, pp. 459–479.

Vu, L.N., Morelli, J.N., Szklaruk, J., 2018. Basic MRI for the liver oncologists and surgeons. J. Hepatocell. Carcinoma 37–50.

Wang, W., Chen, C., Ding, M., Yu, H., Zha, S., Li, J., 2021. Transbts: Multimodal brain tumor segmentation using transformer. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24. Springer, pp. 109–119.

Wang, F., Cheng, C., Cao, W., Wu, Z., Wang, H., Wei, W., Yan, Z., Liu, Z., 2023. MFCNet: A multi-modal fusion and calibration networks for 3D pancreas tumor segmentation on PET-CT images. Comput. Biol. Med. 155, 106657.

Xiao, X., Zhao, J., Qiang, Y., Chong, J., Yang, X., Kazihise, N.G.-F., Chen, B., Li, S., 2019. Radiomics-guided GAN for segmentation of liver tumor without contrast agents. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22. Springer, pp. 237–245.

Xing, Z., Yu, L., Wan, L., Han, T., Zhu, L., 2022. Nestedformer: Nested modality-aware transformer for brain tumor segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part V. Springer, pp. 140–150.

Xu, C., Song, Y., Zhang, D., Bittencourt, L.K., Tirumani, S.H., Li, S., 2023. Spatiotemporal knowledge teacher–student reinforcement learning to detect liver tumors without contrast agents. Med. Image Anal. 90, 102980.

Xu, C., Xu, L., Ohorodnyk, P., Roth, M., Chen, B., Li, S., 2020. Contrast agent-free synthesis and segmentation of ischemic heart disease images using progressive sequential causal GANs. Med. Image Anal. 62, 101668.

Xu, C., Zhang, D., Chong, J., Chen, B., Li, S., 2021. Synthesis of gadolinium-enhanced liver tumors on nonenhanced liver MR images using pixel-level graph reinforcement learning. Med. Image Anal. 69, 101976.

Xu, C., Zhang, D., Song, Y., Kayat Bittencourt, L., Tirumani, S.H., Li, S., 2022. Contrast-free liver tumor detection using ternary knowledge transferred teacher-student deep reinforcement learning. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part V. Springer, pp. 266–275.

Zhang, D., Chen, B., Chong, J., Li, S., 2021. Weakly-supervised teacher-student network for liver tumor segmentation from non-enhanced images. Med. Image Anal. 70, 102005.

Zhang, Y., He, N., Yang, J., Li, Y., Wei, D., Huang, Y., Zhang, Y., He, Z., Zheng, Y., 2022. Mmformer: Multimodal medical transformer for incomplete multimodal learning of brain tumor segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part V. Springer, pp. 107–117.

Zhang, Z., Liu, Q., Wang, Y., 2018. Road extraction by deep residual u-net. IEEE Geosci. Remote Sens. Lett. 15 (5), 749–753.

Zhang, D., Xu, C., Li, S., 2023. Heuristic multi-modal integration framework for liver tumor detection from multi-modal non-enhanced MRIs. Expert Syst. Appl. 221, 119782.

Zhao, J., Li, D., Kassam, Z., Howey, J., Chong, J., Chen, B., Li, S., 2020. Tripartite-GAN: Synthesizing liver contrast-enhanced MRI to improve tumor detection. Med. Image Anal. 63, 101667.

Zhao, J., Li, D., Xiao, X., Accorsi, F., Marshall, H., Cossetto, T., Kim, D., McCarthy, D., Dawson, C., Knezevic, S., et al., 2021. United adversarial learning for liver tumor segmentation and detection of multi-modality non-contrast MRI. Med. Image Anal. 73, 102154.

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2881–2890.

Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J., 2018. Unet++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4. Springer, pp. 3–11.

Zhou, T., Ruan, S., Canu, S., 2019. A review: Deep learning for medical image segmentation using multi-modality fusion. Array 3, 100004.