TAUNet: a triple-attention-based multi-modality MRI fusion U-Net for cardiac pathology segmentation

Dapeng Li¹ · Yanjun Peng^{1,2} · Yanfei Guo¹ · Jindong Sun¹

Received: 5 September 2021 / Accepted: 17 January 2022 / Published online: 8 February 2022
© The Author(s) 2022

Abstract

Automated segmentation of cardiac pathology in MRI plays a significant role for diagnosis and treatment of some cardiac disease. In clinical practice, multi-modality MRI is widely used to improve the cardiac pathology segmentation, because it can provide multiple or complementary information. Recently, deep learning methods have presented implausible performance in multi-modality medical image segmentation. However, how to fuse the underlying multi-modality information effectively to segment the pathology with irregular shapes and small region at random locations, is still a challenge task. In this paper, a triple-attention-based multi-modality MRI fusion U-Net was proposed to learn complex relationship between different modalities and pay more attention on shape information, thus to achieve improved pathology segmentation. First, three independent encoders and one fusion encoder were applied to extract specific and multiple modality features. Secondly, we concatenate the modality feature maps and use the channel attention to fuse specific modal information at every stage of the three dedicate independent encoders, then the three single modality feature maps and channel attention feature maps are together concatenated to the decoder path. Spatial attention was adopted in decoder path to capture the correlation of various positions. Once more, we employ shape attention to focus shape-dependent information. Lastly, the training approach is made efficient by introducing deep supervision mechanism with object contextual representations block to ensure precisely boundary prediction. Our proposed network was evaluated on the public MICCAI 2020 Myocardial pathology segmentation dataset which involves patients suffering from myocardial infarction. Experiments on the dataset with three modalities demonstrate the effectiveness of fusion mode of our model, and attention mechanism can integrate various modality information well. We demonstrated that such a deep learning approach could better fuse complementary information to improve the segmentation performance of cardiac pathology.

Keywords Cardiac pathology MRI · Attention mechanism · Multi-modality fusion · Segmentation

Introduction

Cardiac diseases are of the most dangerous disease causes of death globally. To improve diagnosis and treatment of cardiac diseases, medical imaging techniques such as magnetic resonance imaging (MRI) are now widely used and cardiac MRI is currently regarded as the gold standard for cardiac disease analysis [42]. Myocardial infarction (MI) is one of the most common cardiac diseases that may lead to heart failure. Assessment of myocardial viability is essential in the diagnosis and treatment management for patients suffering from myocardial infarction [55]. Myocardial infarction area refers to the myocardial tissue that may undergo complete ischemic necrosis when myocardial infarction occurs. It consists of two parts: (1) scars, or area of irreversible myocardial injury, histologically manifested as myocardial cell necrosis.

✉ Yanjun Peng
pengyanjuncn@163.com
Dapeng Li
lidapeng636@163.com
Yanfei Guo
201881060069@sdu.edu.cn
Jindong Sun
jalexdo00@gmail.com

¹ College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, Shandong, China

² Shandong Province Key Laboratory of Wisdom Mining Information Technology, Shandong University of Science and Technology, Qingdao 266590, Shandong, China

(2) In the area of reversible myocardial injury, the histological manifestation is edema of myocardial cells. The identification of salvageable myocardium after perfusion and the early restoration of blood flow reperfusion are of great significance in the prognosis of patients with acute myocardial infarction. However, this task remains challenging due to two reasons: (1) myocardial pathology regions vary significantly in the visual appearance. (2) the borders between pathology regions and surrounding normal organs or tissues appear blurry and ambiguous. Cardiac multi-modality MRI is particularly used to provide imaging anatomical and functional information of heart, because it can provide complementary information [29]. The Late Gadolinium Enhancement (LGE) CMR is a T1-weighted, inverted recovery, gradient echo sequence that can enhance infarcted myocardium and has a significant brightness compared to healthy tissue. bSSFP cine CMR is a balanced steady state, freely precessing movie sequence, which can learn heart movement and obtain a clear heart boundary. T2 CMR is a T2-weighted black blood spectrum pre-saturation attenuation inversion recovery sequence that can provide imaging of acute injury and ischemic areas. It can be clearly seen that the bSSFP film CMR sequence has stronger contrast than the other two sequences. For example in Fig. 1, LGE MRI can visualize and detect MI regions, the T2-weighted MRI provides clear visibility images of acute injury and ischemic regions, and the bSSFP MRI offers clear boundaries between anatomical classes and captures cardiac motion. How to fuse the modality specific rich and reliable information to classify the myocardial pathology into nor-

mal, infarct and edema regions is a great valuable subject in clinical diagnose and treatment MI practice. Being aware of the complementary unique features of cardiac multi-modality MRI and the boundary shape attention are helpful for cardiac pathology segmentation.

Conventionally, myocardial pathology segmentation approaches are mainly thresholding-based methods. However, these approaches are easily influenced by the quality of images [32]. Recently, image fusion approaches based on machine learning concept are presented to segment the pathology regions [54]. For instance, Yang used Support Vector Machine (SVM) method to segment the scar on left atrium myocardium [43]. The machine learning-based approaches have improved the segmentation performance in many pathology tasks, for example liver lesion pathology region and brain tumor region [47,48]. At present, motivated by the success of deep learning, convolutional neural networks(CNN) have been extensively used for automatic segmentation of multi-modality medical images. Lots of papers have presented researches with CNN-based methods in multi-modality medical image fields, such as in brain [10,30,36], pancreas [26,52], and multi-organ [7,50]. For cardiac multi-modality field, many deep learning works combining multiple information from specific modality have been reported. Havaei et al. concatenated feature maps extracted from multi-modality into multi-channel, then use CNN layers to fuse the multi-channel information [11]. Liao et al. proposed a multi-modality transfer learning network with adversarial training for 3D multi-modality whole heart

Fig. 1 Cardiac multi-modality MRI. First row, from left to right are bSSFP, LGE, T2-weighted image, respectively. Second row is the Ground Truth of labels, among which green region denotes RV, cyan region denotes LV, blue region denotes normal myocardium, red region denotes edema and white region denotes scar

segmentation [23]. Dolz et al. used cross-modality convolution fusion model to merge the features across network layers [5]. A considerable number of CNN networks have been applied to the multi-modality myocardium pathology segmentation. Elif et al. [6] using a UNet convolution neural network architecture built from residual unit trained by augmentation operations. They only concatenate multi-modality image as channel directly which failed to better fuse the complementary information. Although the dice score of scar+edema is high, the dice score of scar got an undesirable result. Liu et al. [24] proposed a novel fully automatic two-stage method to extract different features of each modality as well as segment myocardium edema and scars. In the first stage, a UNet was trained on bSSFP images to locate the coarse position of the myocardium and obtain the mask of the myocardium as a constraint on the next stage. In the second stage, with the T2 images, LGE images and predicted myocardium masks concatenated as inputs, an M-shaped network based on attention mechanism was trained to segment the myocardial edema and scars accurately. This two-stage method is complex and they also ignore the individual characteristics of the modality only by using channel concatenate to fuse multi-modality images. Yu et al [44] embed a dual-branch attention module in UNet. One of the branches provides channel attention via emphasizing feature association among different channel maps, while the other branch provides spatial attention which adaptively aggregates the features at relative positions regardless of their distances in a weighted manner. Although this model has achieved good results, it lacks the consideration of the fusion of modalities and the focus on boundary information. Jiang et al. [17] present the Max-Fusion U-Net which containing dedicated encoders and spatial attention decoders. However, this model cannot outperform other models in segmenting both scar and scar+edema. In conclusion, lack of the consideration of fuse multi-modality effectively and the focus on the boundary caused the above models which have a inferior effect in multi-modality myocardial pathological segmentation.

As we confirm that the complementary information is of great importance for cardiac pathology segmentation, applying multi-modality images effectively can not only reduce the information uncertainty but also improve clinical diagnosis and segmentation accuracy [11]. However, at present, most DNN-based myocardial pathology segmentation methods are focus on mono-modality CMR, such as LGE. Besides, the existing fusion strategy convolutional networks used for multi-modality MRI segmentation are simple earlier fusion which concatenate multi-modality image as multi-channel directly, they focus on the design of segmentation network architecture, therefore they lack analysis of how to fuse the different feature information and consideration the relationship between multi-modality images with heterogeneous

intensity distributions. Furthermore, as the neural structure that contains the attention mechanism is gradually developing. There are several types of visual attention used in the field of computer vision. The core idea of attention mechanism is to find the correlation between them based on the original data, and then highlight some of its important features, such as channel attention, pixel attention, multi-level attention, etc. In this pathology segmentation work, pay more attention on the position relationship with different anatomical class, is very helpful to segment the pathology with irregular shapes and small region at random locations accurately. But, some works use channel attention or spatial attention to focus on the boundary is not enough.

Motivated by the above observation, in this paper, we propose a fully automatic framework for segmenting cardiac anatomical structures and pathology in multi-modality MRI to tackle the above issues for promoting the segmentation performance. First, to learn complex and complementary feature information of each modality to reduce the information uncertainty, we use later layer-level fusion strategy. The three modality images are used as single input to train individual encoder feature extract network, and these learned modality specific features maps will be fused in every down-sample layer to construct a fusion encoder architecture. The fused representations concatenate three modality specific features and will be fed into the decoder network. Second, for improving the robustness of pathology segmentation performance and the accuracy of clinical diagnosis, channel attention is used in fusion encoder to learn multiple modality information, spatial attention block, channel attention block and dilated convolution blocks (DCB) are used to pay more attention on the position relationship with different anatomical class and extract multi-scale context information in decoder path. Third, we employ shape attention on the three modalities to focus on shape-dependent information, thus to ensure an accurate pathology edge segmentation. For better edge segmentation of myocardial pathology, neither networks have designed a specially boundary attention path before. Our work is the first time that applied a specially shape attention path to pay more attention on boundary information. Lastly, the training approach is made efficient by introducing deep supervision mechanism with object contextual representations (OCR) blocks to make a precise boundary prediction [39,45].

As such, the contributions of this work are summarized as follows:

- We design three dedicated encoders to extract independent specific modal features, and one fusion encoder to learn multiple modality information.
- We use the channel attention in fusion encoder to fuse specific modal information from the three independent encoder, then the fusion information concatenates

independent encoder feature maps into decoder path where channel attention and spatial attention are used to extract the position correlation information and to enhance specify feature maps of different channels.

- A shape attention mechanism is introduced to focus on the segmentation of pathology edge.
- Deep supervision with OCR blocks is adopted in training process to ensure precisely boundary prediction.

We conduct sixfold cross-validation experiment and evaluate our model on the public MICCAI-2020 cardiac multi-modality MRI pathology segmentation challenge (MyoPS 2020) dataset. Experiments demonstrate that our method achieves competitive segment results on this dataset.

The remainder of this paper is organized as follows. In “Related work”, we give a brief of related works in the field of multi-modality fusion pathology segmentation. “Methods and materials” presents our proposed TAUNet architecture. The experiments and results are detailed in “Experiments and results”. Discussion and Ablation are illustrated in “Discussion and ablation”. Finally, the conclusion is presented in “conclusion”, and acknowledgement is presented in last section.

Related works

A group of representative deep learning-based cardiac pathology MRI segmentation methods are proposed in the last decade. And some challenge datasets have been broadly accepted by cardiac MRI segmentation community to gauge performances of the segmentation methods.

Pathology segmentation

Pathology segmentation in MRI plays a vital role for disease diagnosing and treating. Because of irregular shape and small region at random locations, pathology segmentation has always been a challenging task. Liver lesion pathology, brain tumor pathology and myocardial pathology are the common seen fields in automatic medical image pathology segmentation. In liver pathology segmentation, an improved segmentation approach based on watershed algorithm, neurotrophic sets, and fast fuzzy c-mean clustering algorithm is proposed [1]. In brain tumor pathology segmentation, Sun et al. proposed a multi-pathway architecture method based on 3D fully convolutional network with a set of effective training schemes to predict brain tumor pathology [34]. Zhang et al. designed a novel task-structured brain tumor segmentation network, specifically, a modality-aware feature embedding mechanism is introduced to infer the important weights of the modality data during network learning [47]. Myocardial pathology segmentation is also a significant area of pathology

segmentation, fully automated deep learning-based segmentation of normal, infarcted and edema regions from MRI is essential clinical process for patient with MI and other myocardial diseases. Late gadolinium enhancement (LGE) cardiac MRI provides an important protocol to visualize MI. In MICCAI 2020 Myocardial pathology segmentation challenge, Zhai et al. adopted a coarse-to-fine segmentation strategy that contains two segmentation neural networks, the coarse segmentation framework is to predict approximate position of cardiac structures while the fine segmentation framework concatenates output of the coarse network as prior location information to conduct the detailed pathology predictions [46]. They were the winner of the challenge. In their work, multi-modality images are fused channel by channel as the multi-channel inputs to learn a fused feature representation but neglect to fully exploit the multi-modal features. Zhang et al. proposed two cascade network, one is anatomical structure segmentation network, the other is pathological region segmentation network, modality specific feature fused by channel attention block with layer-level fusion strategy [53], this framework achieves good result but requires the image to go through segmentation network twice to get the final result. Zhang et al. trained three parallel segmentation networks, and averaged the prediction with threshold 0.5 in the decision-level [51]. Their method gets a competitive segmentation performance but needs sufficient powerful GPU to train the three parallel model. Martin-Isla et al. used a two-stage method which contains detect network and segment network, the final test dice score were achieved by ensemble of 15 trained models [27]. While awarded as the best paper for their outstanding dice result, their method need complex ensemble network design.

Multi-modality fusion

Multi-modality is widely used in medical image segmentation for its ability to provide multiple and complementary information about target region. Fusing multi-modality information effectively can improve the segmentation performance, so the fusion strategy is worthy of study. According to the multi-modality fusion strategies, Zhou et al. categorized the network architectures into input-level fusion network, layer-level fusion network and decision-level network [54]. The input-level fusion is an earlier fusion strategy, which is very simple and focus on designing network architecture, the multi-modality information is concatenated by channel to learn fused feature. But this strategy lacks of consideration of the specific modality information. Wang et al concatenated the four modalities (T1, T1c, T2, and Flair of MRI) directly by channel, the fused input enter into the segmentation network to segment the brain tumor into whole tumor, tumor core and enhancing tumor core areas [37]. Iqbal et al fused the multi-modal MRI channel by channel, then used a

model which consists of multiple neural network layers connected in sequential order with the feeding of Convolutional feature maps at the peer level to realize the final segmentation [14]. Wang et al. introduced a wide residual network and pyramid pool network (WRN-PPNet), the WRN is used to extract features of multimodal brain tumor slices which are proved to have strong expressive ability, then representation with different level obtained by PPNet is stacked on the features from WRN [41]. With the success of Generative Adversarial Network (GAN), a great number of methods apply the GAN network to improve the segmentation performance. Huang et al. proposed a multi-task coherent modality transferable GAN to address this issue for brain MRI synthesis in an unsupervised manner [12]. Huo et al. employed a GAN-based network to train [13]. In the layer-level fusion method, each modality trains a individual network, and the output individual features will be fused in the layer of the next network, the layer-fusion network can effectively integrate and fully leverage multi-modality information. Chen et al. proposed a novel voxelwise residual network integrating multi-modal and multi-level information, and leveraged auto-context method by integrating image appearance and context for improving performance [3]. Dolz et al. proposed a 3D fully convolutional neural network based on DenseNet, each modality has a path and dense connections to learn more complex feature representation [5]. The decision-level fusion can be seen as a later fusion strategy, later fusion pays more attention on the fusion problem, each modality input train each segmentation network, the output of individual network will be integrated to get the final segmentation result. Nie et al. trained individual modality network and then fused multiple modality features from high-layer of each network [28]. Kamnitsas et al. leveraged three individual network and then average the confidence of each network, the highest confidence voxel is treated as the final segmentation [18]. In MICCAI 2019 Multi-Sequence Cardiac MR (MS-CMR) segmentation challenge, many methods use later layer fusion strategy. Chen et al. formulated an unsupervised learning algorithm by decomposing the training process into two stages, a multi-modal unsupervised image-to-image translation network and a cascade U-Net segmentation network [2]. Wang et al. focused on the alignment of features that are extracted from source (bSSFP and T2) and target (LGE) by a attention U-Net segmentation network with GFRM block [38]. Ly et al. used style data augmentation to prevent the model from over-fitting to any specific contrast and to focus the optimization on the fundamental geometry features of the target [25]. Wang et al. used a supervised method and a modified 2D U-Net equipped with squeeze-and-excitation residual module and the selective kernel module [40]. Li et al. proposed a multiscale dual-path feature aggregation network to solve the shape discontinuity and misclassification problem [21].

Methods and materials

We propose a new cardiac MRI pathology segmentation architecture known as Triple Attentive U-Net (TAUNet) which adopts the shape attention, channel attention and spatial attention. In this model, based on U-Net architecture, we adopt channel attention block after concatenating the three modality features to extract modality channel information adaptively. The spatial attention block is used in decoder path to capture the correlation information of various positions after concatenating skip connection features. Shape attention is introduced to every single modality encoder to detect the pathology edge information.

Our proposed model also uses the dilated convolution and dense connection techniques. In the final layer, owing to the limited cardiac MRI data, we add a deep supervision mechanism with object context representation blocks to learn more representative features by formulating the optimization function aggregating the low-level side output feature maps. This mechanism can improve the training efficiency and generalization performance of the model. Figure 2 presents schematics of the proposed TAUNet, and the structure details of the proposed networks are shown in Table 1.

Independent modality encoder

The three modality cardiac MRI of LGE, T2 and bSSFP have different pixel intensity distribution, thus they contain different modality specific information. For instance, LGE and T2 MRI contain the MI region information, While the bSSFP MRI contains anatomical structural information. Motivated by the above observation, we design three dedicated encoders to extract independent specific modal features to accommodate intensity heterogeneous between modalities. Inspired by U-Net and Tiramisu network [16], we use part of pre-trained DenseNet-121 network as our independent modality encoder which contains dense blocks and transition down blocks. Dense block is composed of Batch Normalization, followed by ReLU activation function, a same convolution with 3×3 and dropout layer with probability $p = 0.2$. Transition down is composed of Batch Normalization, followed by ReLU, a 1×1 convolution, dropout layer with probability $p = 0.2$, and a non-overlapping maxpooling layer with size 2×2 . From the top layer to the bottom layer, the size of feature maps is reduced from 256 to 16, while the channel number of feature map is increased to 1024. Each modality original input image size is 256×256 , we expand one dimension and then concatenate three original images as three channels in data preprocessing stage. Then the each modality image with size of $256 \times 256 \times 3$ is used as input into independent modality encoder.

Fig. 2 The proposed TAUNet. The main architecture is similar to the U-Net structure; however, the encoder path is divide into independent modality encoder path and fusion modality encoder path to extract different modality information. We also add a shape attention path to learn edge information

Table 1 Parameter setting of the proposed model

Layer	Filter	Stride	Output shape
Independent encoder 1	64	2	$64 \times 128 \times 128$
Independent encoder 2	128	2	$128 \times 64 \times 64$
Independent encoder 3	256	2	$256 \times 32 \times 32$
Independent encoder 4	512	2	$512 \times 16 \times 16$
Independent encoder 5	1024	2	$1024 \times 8 \times 8$
Fusion encoder 1	64	2	$64 \times 128 \times 128$
Fusion encoder 2	128	2	$128 \times 64 \times 64$
Fusion encoder 3	256	2	$256 \times 32 \times 32$
Fusion encoder 4	512	2	$512 \times 16 \times 16$
Fusion encoder 5	1024	2	$1024 \times 8 \times 8$
BottleBlock	512	2	$512 \times 8 \times 8$
Shape gatedconv 1	32	1	$32 \times 256 \times 256$
Shape gatedconv 2	16	1	$16 \times 256 \times 256$
Shape gatedconv 3	8	1	$8 \times 256 \times 256$
Decoder 1	512	2	$256 \times 16 \times 16$
Decoder 2	256	2	$128 \times 32 \times 32$
Decoder 3	128	2	$64 \times 64 \times 64$
Decoder 4	64	2	$32 \times 128 \times 128$
Decoder 5	32×2	2	$32 \times 256 \times 256$
OCR 1	32	2	$6 \times 256 \times 256$
OCR 2	64	2	$6 \times 256 \times 256$
OCR 3	128	2	$6 \times 256 \times 256$
Output	6×3	2	$6 \times 256 \times 256$

Multi-modality fusion encoder

The architecture of multi-modality fusion encoder path is very much like the independent modality encoder, which also has five times down-sampling. First, we use direct channel concatenation as fusion strategy to concatenate three modality feature maps which derived from every stage of the independent modality encoder. This fusion strategy can maximally keep the original image information, learn the intrinsic image feature and fully leverage multi-modal images. Followed by the above multi-channel concatenation, a channel attention blocks applied for adaptively weighted feature maps of different modality images. The attention block is composed of continuous channel attention, 3×3 2D convolution with padding size 1 and Batch Normalization. Figure 3

Fig. 3 Inner structure of the channel attention block schematic diagram. Each line of it calculates the dependency relationship between channel and other channels, and the value is probabilized by softmax to between 0 and 1. The larger the value, the stronger the dependency. C, H and W in the figure represent channel, height and width, respectively

presents the schematic structure of channel attention. The input feature image A is first reshaped to $(C \times N)$, and then the transposition of A and A is matrix multiplied. After softmax, the map X ($C \times C$) between the channels is obtained, and then multiply $A(C \times N)$. Finally, after adding to the original image, the final output E is obtained. 2D Maxpooling operation is used after the attention block. The fused features extracted from each independent modality encoder stage propagate in the fusion encoder path, such approach is significant in the field of cardiac pathology segmentation, where specific modality imaging focuses on specific pathology. We employ Dense Atrous Convolution (DAC) module, Residual Multi-kernel pooling (RMP) module and 1×1 convolution as our bottleneck [9]. The DAC module is similar to the Inception structure which uses multi-scale atrous convolution to extract rich feature representations. There are four cascaded branches in DAC module, as the number of convolutions gradually increases, from 1 to 1, 3 and 5, and then the receptive field of each branch will be 3, 7, 9, 19. In each atrous branch, we apply a 1×1 convolution for rectified linear activation. The RMP module gathers context information with four pooling kernels with different sizes. Then features are fed into 1×1 convolution to reduce the dimension of feature maps. Finally, the upsampled features are concatenated with original features. Using these two modules can obtain more abstract features and retain more spatial information to improve the performance of medical image segmentation [10]. The two modules follow one 1×1 convolution with Batch Normalization and ReLU operation, which aims to reduce the number of channels. By using the multi-modality fusion encoder path and bottle block, our model can capture multi-scale and multi-modality information to improve pathology segmentation performance.

Decoder with dual attention

In the decoder path of our proposed TAUNet, there are three important modules, respectively, are the cascade dilated convolution block (DCB), dual attention module [8] which contains the spatial attention block and the channel attention block, the SKconv unit [22]. The cascade dilation convolution block is equipped in the concatenate approach, which fuses features from each resolution by using dilated convolutions with different rates and concatenating all convolution outputs as the input into the next block. There are five 3×3 kernel with dilate rate 1, 3, 6, 9, 12 branches in our cascade DCB module. We employ residual units instead of the plain units, to accelerate the training and convergence. Each residual unit contains two SKconv units, each of which is followed by a batch normalization layer and rectified linear unit (ReLU) activation. The SKconv unit can adaptively adjust the size of the respective field to capture more informative multi-scale spatial features using kernel size of 3

and 5. Cascade dilation convolution block and SKconv unit can improve the ability to obtain receptive fields of multiple size and to learn multi-scale features. Since pathology region with irregular shape often takes up very small space of the whole MRI, the class imbalance problem is worth considering. To tackle this issue, we embed dual attention module which contains spatial attention block and channel attention block. Spatial attention block can help to capture class position information and distribute different importance on different classes, while channel attention enhances specific feature maps among different channels. This module effectively enhances the feature representation without adding too many parameters. The channel attention block and spatial attention block are depicted in Figs. 3 and 4. To make full use of long-range context information, the feature maps of these two attention blocks are aggregated. The specific process is described as follows. first, The previous feature is sent to the spatial attention module and the channel attention module separately after dimensionality reduction by a 1×1 convolution. At this time, the number of channels of the feature becomes a quarter of the previous original feature. Then, the output of the two attention blocks are converted through the convolutional layer, batchnorm layer and ReLU operation to output the specified number of channels, respectively. An element-wise summation is performed to achieve dual attention feature fusion. Finally, a 1×1 kernel size convolution with dropout layer is followed to obtain the final predicted feature map. As for the spatial attention block, the input local feature image $A(C \times H \times W)$, we first use the convolution operation to obtain three feature maps of B , C , and D , and reshape them to $(C \times N)$, respectively. After multiplying the transpose of B and C , the spatial attention map $S(N \times N)$ is obtained through softmax. Next, multiply the transpose of S by D , and reshape the result to $(C \times H \times W)$. Finally, the result obtained above is multiplied by the scale factor and then the original image is added to obtain the output map E .

Shape attention path

In addition to the above U-Net likes convolution neural network (CNN), we add an auxiliary stream called shape attention path to learn shape and boundary information. Image segmentation dataset includes color, shape and texture information, but the existing network architecture is to input all information directly to the network, such processing is not the optimal choice. Therefore, Sun et al. proposed a two-stream CNN to process different information [33]. Two-stream includes regular CNN stream and shape stream which deal with different emphases, shape stream is dedicated to deals with information related to the boundary, and gated convolutional layer (GCL) is a core component of this architecture that helps to process information only related to the boundary and filter out the rest. First step of the specific

Fig. 4 Inner structure of the decoder block schematic diagram. Cascade DCB and SKConv are the main components to aggregate multi-scale features and adjust the size of receptive field adaptively. We embed the dual attention module [8] which contains spatial attention block

and channel attention block to capture the global context information on feature maps and enhance specific feature maps among different channels

process is to join the two stream outputs, then use 1×1 convolution kernel and sigmoid function to get an attention distribution, and through a GCL layer to get a new shape distribution finally. The process is illustrated as follow in Eqs. (1) and (2), where $C_{1 \times 1}$ denotes normalized 1×1 convolution applied on feature map, r denotes the residual block applied on feature map. t is the layer number of our shape stream, (i, j) denotes a point of feature map. GCL outputs an attention map α_t , s_t denotes shape stream feature map. σ is the sigmoid function, and ω is the weight hyper-parameter of GCL. Given the attention map α_t , GCL is applied on s_t as an element-wise product $\odot$ with attention map α_t followed by residual connection and channel-wise weighting with kernel ω_t . At each pixel (i, j) , GCL $\otimes$ is computed as Eq. (2). $\hat{s}_t$ is then passed on to the next layer in the shape stream for further processing. The two streams are parallel and cooperate with each other.

Motivated by this paper, we propose shape attention path on three modality MRI to learn the pathology shapes. The shape attention path is presented in Fig. 5.

A edge loss is supervised by this path, which is the binary cross-entropy loss between the shape attention block output and the mask class boundary. Since the information displayed by each modality is different, we integrate the edge infor-

mation of the three modality MRI to improve the overall boundary segmentation performance.

$$\alpha_t = \sigma(C_{1 \times 1}(s_t \parallel r_t)), \quad (1)$$

$$\begin{aligned} \hat{s}_t^{(i,j)} &= (s_t \otimes w_t)_{(i,j)} \\ &= \left(\left(s_{t(i,j)} \odot \alpha_{t(i,j)} \right) + s_{t(i,j)} \right)^T w_t. \end{aligned} \quad (2)$$

Deep supervision with OCR

Isensee et al. had concluded that deep supervision ensembles different resolutions feature maps by applying supervision on small size intermediate inputs to have a better prediction [15], so we do the same way. In the feature restoring part, we ensemble the side output from low-level feature maps to make the final prediction. We apply the OCR block on the side output of each level feature map, and then resample the smaller-sized feature map to the same size as the top-level feature map using the mode bilinear interpolate.

OCR (illustrated in Fig. 6) is a context aggregation strategy which uses region class to refine pixel class. First,

Fig. 5 Inner structure of shape attention path schematic diagram. The gated convolutional layers are used to fuse texture and shape information, while residual layers are used to fine-tune the shape features [35]

Fig. 6 Inner structure of the Object Contextual Representation block schematic diagram [45]. We merge the output of the Augmented Representations and Soft Region Representations by using the way of channel concatenate as the final output and then ensemble to the deep supervision mechanism

the target region is obtained by supervised learning of the groundtruth, and use 3x3 convolution and other operations to obtain the overall pixel representations. K is the number of classes, K feature maps are output from the network, each represents a segmentation map of a class, and groundtruth is used to supervise the classification during training. This step is the result of calculating a coarse segmentation, which is the soft object region. Second, by aggregating K soft object regions (K soft object regions) and overall pixel representations, the object region representations (Object Region Representations) are obtained. Finally, we aggregate K target regions and consider their relationship with all pixels to enhance the representation of each pixel. Eventually we will get the Object Contextual Representations.

Motivated by the observation, we leverage deep supervision to combine with OCR to make a better and robust prediction. We made a small change to the original OCR that we merge the output of the Augmented Representations and Soft Region Representations using the way of channel concatenate as the final output to realize the deep supervision.

Loss function

We propose an objective function optimizes for precise segmentation and urge the network to learn the boundary information. In our multi-modality cardiac MRI pathology segmentation task, there exists a highly class imbalance problem and irregular shape boundary problem. To overcome the class imbalance between the region of interest structures and surrounding background, we employ a loss function to incorporate both the dice loss and cross-entropy loss. The cross-entropy loss can optimize the voxel-level accuracy, whereas the dice loss helps to improve the segmentation metrics. We define the dice loss and cross-entropy loss as in Eqs. (3) and (4), where Ω denotes the domain of all pixels, y and $\hat{y}$ are the ground truth one-hot class encodes matrix and the predicted probabilities matrix of each pixel, respectively. K is number of label classes and y_i^k denotes i^{th} pixel of the k^{th} indexed label classes of matrix y , y_{ic} is a symbolic function, if the true category of the sample i is equal to c , then y_{ic} is 1, otherwise it is 0, P_{ic} is predicted probability of the

observation sample i belonging to the category c . In Eq. (4), after the output vector of the model passes through the softmax activation function, the vector is normalized into a probability form, and then calculated with the target to calculate the cross-entropy loss in the strict sense. The target refers to the ground truth one-hot matrix which encoding the ground truth class of each pixel. Previous classic paper has leveraged the above two loss as a combination loss function to train the cardiac segmentation model, and got amazing dice coefficient result [19]. Motivated by this, we also adapt the combination loss as part of our loss function. Since dice loss puts more emphasis on the overall similarity coefficient, we empirically set weight $\lambda_1 = 1$ and $\lambda_2 = 1.5$ to each of the two loss functions. To learn boundary information, we employ edge loss derived from shape attention module to denote the binary cross entropy loss of the predicted shape boundary. We set $\lambda_3 = 1$ as the weight of edge loss. In our experiments, we found that the weights setting works well. The overall loss function can be seen in Eq. (5).

$$L_{Dice}(\hat{y}, y) = 1 - \frac{2}{K} \sum_{k=0}^{K-1} \frac{\sum_i y^k \hat{y}_i^k}{\sum_i y^k + \hat{y}_i^k}, \quad (3)$$

$$L_{CE} = -\frac{1}{|N|} \sum_i \sum_{c=1}^K y_{ic} \log(P_{ic}), \quad (4)$$

$$L_{total} = \lambda_1 L_{CE} + \lambda_2 L_{Dice} + \lambda_3 L_{Edge}. \quad (5)$$

Experiments and results

For this study, we implement the model using PyTorch and conduct our experiments on the MyoPS 2020 segmentation dataset. We present our results in the following sections. The experiments were carried out on one NVIDIA RTX 2080TI GPU.

Data and preprocess

The dataset provide 45 cases of multi-modality cardiac MRI, each case refers to a patient with three modality cardiac MRI, those being the LGE, T2 and bSSFP cardiac MRI. All these clinical data have got institutional ethic approval and have been anonymized. The data released here have been registered using the MvMM method [55], three modality cardiac MRI have been aligned into a common space and to resampled into same spatial resolution. The ground truth provides labels including: left ventricle (LV), right ventricle (RV), LV normal myocardium, LV myocardial edema, LV myocardial scar, though the evaluation of the test data will focus on the myocardial pathology segmentation. We train our model to segment both anatomical structures and

pathology on training dataset, then test the effectiveness of our model in segmenting pathology on testing dataset. The ground truth of testing dataset had been encrypted; therefore, we can only get the dice results between our predicts and the encrypted ground truth.

We conduct sixfold cross-validation experiment on the training dataset. Each image slice was z score normalized based on the mean and standard deviation value. The following data augmentations were performed during runtime: random rotations transformation between -90 and 90 degree, random horizontal and vertical flips transformation with chance 50%, lastic deformations transformation, and gamma shifts transformation with the scope of $0.5 \sim 1.5$. Contrast Limited Adaptive Histogram Equalization (CLAHE) is applied to each training image slices to weaken the intensity inhomogeneity problem. A center-crop of resolution 256 px by 256 px was made, and zero-padding was applied if necessary. The model was trained using RAdam for 500 epochs with $\beta_1 = 0.9$ and $\beta_2 = 0.999$, along with weight decay value of $1E-4$ and initial learning rate of $5E-4$ exponentially decayed with parameter 0.99. Two-dimensional batch normalization was used with a batch size of 4.

Postprocess

In the postprocessing stage, morphological anatomical constraints is used to refine the results outputted by our TAUNet. Morphological functions of the python skimage library are adopted to implement the morphological postprocessing operations. We found in the ground truth of training dataset that scar and edema region is generally a whole connected area and looks like a partial ring. Motivated by the above observations, we first use closed operation to join adjacent areas that are not connected. Then, we remove the small unconnected areas under a predefined threshold and fill them with adjacent class pixels by using morphological functions.

Results

We present our results on the dataset. The Dice Coefficient metric is used for evaluation consistent with other benchmarks and works. Given two countable sets A and B , the Dice coefficient is formally defined as Eq. (6). Evidently, $Dice(A, B)$ is maximized at 1 when $A = B$ and is minimized at 0 when $A \neq B$.

$$Dice(A, B) = \frac{2|A \cap B|}{|A \cap B| + |A \cup B|}. \quad (6)$$

As a measure of shape similarity, the Hausdorff distance which measures the maximum distance from one group to the nearest point in another group can make a better supplement for Dice. The Hausdorff distance coefficient is defined as

Table 2 Dice scores for MyoPS 2020 segmentation challenge on the cross-validation dataset

Model	LV	RV	Normal myocardium	Scar	Edema
Ding [4]	0.896	0.867	0.860	0.493	0.580
Sinha [31]	0.911	0.883	0.870	0.510	0.605
Jiang [17]	0.930	0.907	0.890	0.608	0.630
Li [20]	0.940	0.913	0.904	0.641	0.752
Yu [44]	0.942	0.917	0.911	0.666	0.778
Ours (TAUNet)	0.940	0.915	0.912	0.627	0.780

Table 3 Hausdorff distance (mm) for MyoPS 2020 segmentation challenge on the cross-validation dataset

Model	LV	RV	Normal myocardium	Scar	Edema
Ding [4]	10.72	11.74	20.30	24.17	32.18
Sinha [31]	9.68	10.55	19.93	20.08	31.09
Jiang [17]	7.59	7.98	18.10	18.03	21.14
Li [20]	7.15	7.53	17.04	17.68	20.87
Yu [44]	6.89	7.00	16.17	16.33	19.75
Ours (TAUNet)	6.95	7.08	16.15	17.83	19.79

Eq. (7): where $h(A, B)$ represents the distance from point A to point B . Correspondingly, $h(B, A)$ represents the distance from B to A .

$$Hausdorff(A, B) = \max[h(A, B), h(B, A)]. \quad (7)$$

We train our proposed model for 500 epochs in 12 h, then we use our model to predict the testing dataset. The Dice scores and Hausdorff distance for the LV, RV, normal myocardium and pathology segmentation achieved on training dataset are presented in Tables 2 and 3. We compare our model with other three models proposed by the challenge participants [17,20,44]. We also use two other state-of-the-art advanced classical multi-modality MRI segmentation models that are proposed by recent papers [4,31] for comparison. Sinha [31] proposed Multi-scale self-guided attention to improve the expression ability of extracted features in brain tumor segmentation. Ding [4] used a multi-path adaptive fusion network which applied skip connection in ResNet to dense block to effectively segment the brain tumor.

Since the parameters of the initialization and convolution network of these models are ambiguous in their original papers, and the fusion methods of different modality are different. For quantitative study, we make some changes in reorganize these models, all the 3D convolution kernels are replaced by 2D kernels to adapt to the pathological segmentation and the hyper parameters are modified to consistent with our TAUNet. As for the fusion method, we directly concatenate three modality as channel dimension to mimic the models proposed by Sinha and Ding. The post-processing and pre-processing methods are the same that to make our comparison to focus on the architecture of model. All the models were trained 500 epochs with the same initial con-

figuration. It can be observed in Table 2 that, all the models achieve good dice performance in anatomical structural segmentation (e.g., LV, RV, Normal myocardium), but in terms of pathological segmentation (e.g., Edema and Scar), the performance of these methods are different. Yu et al. outperform other models in segmenting RV and LV, while our TAUNet shows the best performance in segmenting normal myocardium, both of our two models are superior to other models in segmenting scar.

These experiments demonstrate that our triple attention modality fusion network is effective in both anatomical structural and pathological segmentation. Since our model integrates multi-path encoder, shape path and spatial attentive decoder, the complexity has increased accordingly. It will spend more time on training and predicting approach compared with other classical models.

During the testing stage, we perform the test time augmentation (TTA) techniques by mirror flipping and rotating, then averaging the outputs of the two augmentation modes and the original input. Lastly we sent the prediction onto the evaluation system. As for the testing dataset, the mean dice results on the evaluation system are presented in Table 4, compared with state-of-the-art methods. The evaluation system only provides pathology segmentation results, i.e. scar dice coefficient, scar+edema dice coefficient. The prediction time on 20 testing patient data takes about 1.5 min. Figure 7 presents our segmentation results on the testing dataset compared with Five other conventional models. The visualized predictions showed that our method output look more like the actually shape of pathology. It is reported that the inter-observer variation of manual scar segmentation, in terms of Dice overlap, was 0.5234 ± 0.1578 , so deep learning approach is superior to the judgement of clinicians.

Table 4 Comparison with state-of-the-art methods of the dice score on testing dataset evaluate system

Model	Scar Dice	Scar+Edema Dice
Zhang [49]	0.647	0.709
Martin-Isla [27]	0.665±0.143	0.698±0.128
Liu [24]	0.638 ± 0.097	0.561 ± 0.103
Yu [44]	0.634 ± 0.289	0.688 ± 0.148
Li [20]	0.605 ± 0.263	0.655 ± 0.137
Jiang [17]	0.584 ± 0.263	0.614 ± 0.178
Dolz [5]	0.568 ± 0.232	0.611 ± 0.111
Zhang [49]	0.468 ± 0.268	0.557 ± 0.183
Sinha [31]	0.459 ± 0.255	0.593 ± 0.268
Elif [6]	0.448 ± 0.173	0.682 ± 0.163
Ding [4]	0.425 ± 0.278	0.542 ± 0.246
Ours (TAUNet)	0.636 ± 0.239	0.695 ± 0.136

Our segmentation results, both the mean and the standard deviation coefficient of dice are competitive

Fig. 7 Segmentation results on MyoPS 2020 testing dataset. For the left of the arrow, the columns from left to right indicate the three modality MRI. For the right of the arrow, the columns from left to the right indicate predictions of our model, other four classical modality fusion

models, respectively. The rows from top to bottom indicate image slices from apex to basal. In the prediction images, brown region denotes scar, and green region denotes edema

In the MyoPS 2020 challenge, Martin-Isla et al. was awarded as the best paper which used a two-stage method containing detect network and segment network, the final test dice score was achieved by ensemble of 15 trained models. Zhang et al. [49] showed the best segmentation performance, they used EfficientNet as encoder backbone to improve the representation ability and employed a weighted bi-directional feature pyramid network as the decoder to fuse multi-scale feature. The above methods achieve the best segmentation performance but they also have certain disadvantages. They use either multi stage approach or multi model ensemble technique, which are hardware intensive and time inefficient. Our TAUNet uses a one stage end to end segmentation network and achieves a competitive result, especially the dice coefficient of scar+edema pathology which ranking the third place.

Discussion and ablation

To verify the effectiveness of our proposed multi modality fusion method, we performed ablation experiments on the MyoPS 2020 training dataset and testing dataset. We randomly selected 10 data from testing dataset as our validation set A, and also randomly selected 5 data from training dataset as our validation set B. All the ablation models were trained 500 epoches on training dataset, then we used the trained ablation models to make predictions on our two validation datasets. The predictions of validation set A were submitted to the evaluation system for evaluation. To test the effectiveness of the component blocks employed in our model, we use the U-Net with three independent modality encoders as our baseline model, then we add fusion encoder block, channel attention and spatial attention blocks, shape

Table 5 Comparison with ablation methods on validation dataset A

Model	Fusion encoder	Channel and Spatial attention	Shape attention	OCR supervision	Scar	Scar+ Edema
Baseline	×	×	×	×	0.558	0.689
Model 1	✓	×	×	×	0.572	0.698
Model 2	✓	✓	×	×	0.623	0.722
Model 3	✓	✓	✓	×	0.641	0.739
TAUNet	✓	✓	✓	✓	0.672	0.781

Table present average Dice scores of baseline model and three variations of our model. ✓ denotes the corresponding block is used while × denotes block is not adopted

Fig. 8 The visualized statistical dice results of ablation methods on validation dataset A. Left is the statistical results of scar segmentation dice. Right is the statistical results of scar+edema segmentation dice

attention block and deep supervision OCR blocks in turn to conduct ablation studies on the MyoPS 2020 cardiac pathology segmentation dataset. To ensure a fair procedure, all these aforementioned networks share the same hyper-parameter configurations (i.e., initialization weight and bias, and CNN kernel size) as our TAUNet, each model was trained for 500 epochs with a batch size of 4 and supervised by our combination loss function. We use the same channel numbers, pooling manner, and sampling manner. Table 5, Figs. 8 and 9 present comparisons of the ablation studies on validation dataset A while Tables 6, 7 and Fig. 10 present comparisons on validation dataset B. From the view of dice metric in Tables 5 and 6, it can be observed that adopting the fusion encoder, channel attention and spatial attention, and shape attention modules is an effective method to improve the performance of cardiac MRI pathology segmentation. The box plot results in Fig. 8 show that our TAUNet is also a robust method in segmenting scar and scar+edema. In Tables 5 and 6, compared with Model 1 and Model 2, what is worth mentioning is that, the dice coefficient of scar+edema is obviously improved using shape attention block. From the view of

baseline, Model 1 and Model 2, dice coefficient of scar increases step by step. Many methods failed to achieve competitive segmentation results in segmenting the scar region, because the scar region suffers large shape variations. Our proposed network leverages fusion encoder modules to learn more multiple modality information and attention module to extract informative features, thereby significantly improving the pathology segmentation performance. Figure 9 visually compares the cardiac pathological segmentation results on the validation dataset A, while Fig. 10 visually compares the cardiac pathological segmentation results on validation dataset B. The visual comparison between columns shows that U-Net likes model equipped with our proposed component blocks generates a more accurate segmentation boundary, which is especially obvious in the images near apex. In addition, compared with other models, TAUNet achieves the best segmentation performance which are seem to most closely resemble the correct results. From the view of Hausdorff distance metric in Table 7, the effectiveness of each component is consistent with the dice coefficient metric.

Fig. 9 Segmentation results of ablation study on MyoPS 2020 validation dataset A. The rows from top to bottom indicate the image slices from basal to apex. The first three columns from left to right indicate the three input modality MRI. The last columns from left to the right

indicate predictions of the models corresponding to Table 5. In the prediction images, brown region denotes scar, and green region denotes edema

Table 6 Comparison with ablation methods on validation dataset B

Model	Fusion encoder	Channel and Spatial attention	Shape attention	OCR supervision	Scar	Edema
Baseline	×	×	×	×	0.785	0.827
Model 1	✓	×	×	×	0.790	0.868
Model 2	✓	✓	×	×	0.818	0.891
Model 3	✓	✓	✓	×	0.889	0.898
MFAUNet	✓	✓	✓	✓	0.893	0.912

Table present average Dice scores of baseline model and three variations of our model. ✓ denotes the corresponding block is used while × denotes block is not adopted

Table 7 Hausdorff distance (mm) for MyoPS 2020 segmentation challenge on the validation dataset B

Model	Fusion encoder	Channel and Spatial attention	Shape attention	OCR supervision	Scar	Edema
Baseline	×	×	×	×	18.13	17.69
Model 1	✓	×	×	×	17.98	17.56
Model 2	✓	✓	×	×	17.75	17.18
Model 3	✓	✓	✓	×	17.21	16.77
MFAUNet	✓	✓	✓	✓	16.79	15.18

✓ denotes the corresponding block is used while × denotes block is not adopted

Fig. 10 Segmentation results of ablation study on MyoPS 2020 validation dataset B. The rows from top to bottom indicate the image slices from basal to apex. The first three columns from left to right indicate the three input modality MRI. The last columns from left to the right

indicate predictions of the models corresponding to Table 6. In the prediction images, yellow region denotes scar, and green region denotes edema

Conclusion

Automatic segmentation of multi-modality cardiac MRI is significant for the diagnosis of myocardium pathology. In this paper, we propose a new and robust automatic cardiac MRI pathology deep learning segmentation network which is called Triple Attention U-Net (TAUNet). Based on U-Net network, some effective techniques were applied to improve cardiac MRI segmentation performance. We use three dedicated encoders to extract independent specific modal features, then concatenate the modality feature maps and use the channel attention to fuse specific modal information in every stage of the three dedicate encoders. We adopt spatial attention block after concatenation of low-level and high-level features in decoder path, by weighting spatial adaptively, the network has the ability of aggregating different scale features and determining which and where of the scale features should we focus on. Meanwhile a shape attention path is introduced to focus on the segmentation of pathology edge. Finally, deep supervision with OCR blocks is adopted in training process to ensure precisely boundary

prediction. Experiments show that our method achieves competitive results on the dataset, and confirm the robustness and generalizability of the proposed network.

With this work, we hope to take a small step towards the clinical application of deep learning methods, and hopefully inspire more work on pathological segmentation in the future.

Declarations

Funding This work was supported in part by the National Natural Science Foundation of China (Grant no. 61976126), Shandong Nature Science Foundation of China (nos. ZR2019MF003, ZR2017MF054, ZR2020MH132, ZR2020MF291).

Conflict of interest The authors declare there are no conflicts of interest regarding the publication of this paper.

Availability of data and materials Data related to the current study are available from the corresponding author on reasonable request.

Code availability The codes used during the study are available from the corresponding author by request.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Anter AM, Hassenian AE (2019) Ct liver tumor segmentation hybrid approach using neutrosophic sets, fast fuzzy c-means and adaptive watershed algorithm. *Artif Intell Med* 97:105–117
- Chen C, Ouyang C, Tarroni G, Schlemper J, Qiu H, Bai W, Rueckert D (2019) Unsupervised multi-modal style transfer for cardiac MR segmentation. In: *International workshop on statistical atlases and computational models of the heart*. Springer, pp 209–219
- Chen H, Dou Q, Yu L, Qin J, Heng PA (2018) Voxresnet: deep voxelwise residual networks for brain segmentation from 3d MR images. *NeuroImage* 170:446–455
- Ding Y, Gong L, Zhang M, Li C, Qin Z (2020) A multi-path adaptive fusion network for multimodal brain tumor segmentation. *Neurocomputing* 412:19–30
- Dolz J, Gopinath K, Yuan J, Lombaert H, Desrosiers C, Ayed IB (2018) Hyperdense-net: a hyper-densely connected CNN for multi-modal image segmentation. *IEEE Trans Med Imaging* 38(5):1116–1126
- Elif A, Ilkay O (2020) Accurate myocardial pathology segmentation with residual u-net. *Myocardial pathology segmentation combining multi-sequence CMR challenge*. Springer, Berlin, pp 128–137
- Fang X, Yan P (2020) Multi-organ segmentation over partially labeled datasets with multi-scale feature abstraction. *IEEE Trans Med Imaging* 39(11):3619–3629
- 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
- Gu Z, Cheng J, Fu H, Zhou K, Hao H, Zhao Y, Zhang T, Gao S, Liu J (2019) Ce-net: context encoder network for 2d medical image segmentation. *IEEE Trans Med Imaging* 38(10):2281–2292
- Guo Z, Li X, Huang H, Guo N, Li Q (2019) Deep learning-based image segmentation on multimodal medical imaging. *IEEE Trans Radiat Plasma Med Sci* 3(2):162–169
- Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, Pal C, Jodoin PM, Larochelle H (2017) Brain tumor segmentation with deep neural networks. *Med Image Anal* 35:18–31
- Huang Y, Zheng F, Cong R, Huang W, Scott MR, Shao L (2020) Mcm-tan: multi-task coherent modality transferable gan for 3d brain image synthesis. *IEEE Trans Image Process* 29:8187–8198
- Huo Y, Xu Z, Bao S, Bermudez C, Moon H, Parvathaneni P, Moyo TK, Savona MR, Assad A, Abramson RG et al (2018) Splenomegaly segmentation on multi-modal MRI using deep convolutional networks. *IEEE Trans Med Imaging* 38(5):1185–1196
- Iqbal S, Ghani MU, Saba T, Rehman A (2018) Brain tumor segmentation in multi-spectral MRI using convolutional neural networks (CNN). *Microsc Res Tech* 81(4):419–427
- Isensee F, Jaeger PF, Full PM, Wolf I, Engelhardt S, Maier-Hein KH (2017) Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features. In: *International workshop on statistical atlases and computational models of the heart*. Springer, pp 120–129
- Jégou S, Drozdal M, Vazquez D, Romero A, Bengio Y (2017) The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp 11–19
- Jiang H, Wang C, Chatsias A, Tsiftaris SA (2020) Max-fusion u-net for multi-modal pathology segmentation with attention and dynamic resampling. *Myocardial pathology segmentation combining multi-sequence CMR challenge*. Springer, Berlin, pp 68–81
- Kamnitsas K, Bai W, Ferrante E, McDonagh S, Sinclair M, Pawlowski N, Rajchl M, Lee M, Kainz B, Rueckert D, et al. (2017) Ensembles of multiple models and architectures for robust brain tumour segmentation. In: *International MICCAI brainlesion workshop*. Springer, pp 450–462
- Khened M, Kollerathu VA, Krishnamurthi G (2019) Fully convolutional multi-scale residual densenets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. *Med Image Anal* 51:21–45
- Li F, Li W (2020) Dual-path feature aggregation network combined multi-layer fusion for myocardial pathology segmentation with multi-sequence cardiac mr. *Myocardial pathology segmentation combining multi-sequence CMR challenge*. Springer, Berlin, pp 146–158
- Li F, Li W, Qin S, Wang L (2021) Mdfa-net: multiscale dual-path feature aggregation network for cardiac segmentation on multi-sequence cardiac MR. *Knowl-Based Syst* 215:106776
- Li X, Wang W, Hu X, Yang J (2019) Selective kernel networks. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp 510–519
- Liao X, Qian Y, Chen Y, Xiong X, Wang Q, Heng PA (2020) Mmtlnet: multi-modality transfer learning network with adversarial training for 3d whole heart segmentation. *Comput Med Imaging Graph* 85:101785
- Liu Y, Zhang M, Zhan Q, Gu D, Liu G (2020) Two-stage method for segmentation of the myocardial scars and edema on multi-sequence cardiac magnetic resonance. *Myocardial pathology segmentation combining multi-sequence CMR challenge*. Springer, Berlin, pp 26–36
- Ly B, Cochet H, Sermesant M (2019) Style data augmentation for robust segmentation of multi-modality cardiac MRI. In: *International workshop on statistical atlases and computational models of the heart*. Springer, pp 197–208
- Man Y, Huang Y, Feng J, Li X, Wu F (2019) Deep q learning driven CT pancreas segmentation with geometry-aware u-net. *IEEE Trans Med Imaging* 38(8):1971–1980
- Martín-Isla C, Asadi-Aghbolaghi M, Gkontra P, Campello VM, Escalera S, Lekadir K (2020) Stacked BCDU-net with semantic CMR synthesis: application to myocardial pathology segmentation challenge. *Myocardial pathology segmentation combining multi-sequence CMR challenge*. Springer, Berlin, pp 1–16
- Nie D, Wang L, Gao Y, Shen D (2016) Fully convolutional networks for multi-modality isointense infant brain image segmentation. In: *2016 IEEE 13th international symposium on biomedical imaging (ISBI)*, IEEE, pp 1342–1345
- Pei C, Wu F, Huang L, Zhuang X (2021) Disentangle domain features for cross-modality cardiac image segmentation. *Med Image Anal* 20:102078
- Rahmat R, Saednia K, Khani MRHH, Rahmati M, Jena R, Price SJ (2020) Multi-scale segmentation in GBM treatment using diffusion tensor imaging. *Comput Biol Med* 123:103815
- Sinha A, Dolz J (2020) Multi-scale self-guided attention for medical image segmentation. *IEEE J Biomed Health Inform* 20:20
- Spiewak M, Malek LA, Misko J, Chojnowska L, Milosz B, Kłopotowski M, Petryka J, Dabrowski M, Kepka C, Ruzyllo W (2010)

- Comparison of different quantification methods of late gadolinium enhancement in patients with hypertrophic cardiomyopathy. *Eur J Radiol* 74(3):e149–e153
33. Sun J, Darbehani F, Zaidi M, Wang B (2020) Saunet: shape attentive u-net for interpretable medical image segmentation. In: *International conference on medical image computing and computer-assisted intervention*. Springer, pp 797–806
 34. Sun J, Peng Y, Guo Y, Li D (2021) Segmentation of the multimodal brain tumor image used the multi-pathway architecture method based on 3d FCN. *Neurocomputing* 423:34–45
 35. Takikawa T, Acuna D, Jampani V, Fidler S (2019) Gated-scnn: Gated shape CNNs for semantic segmentation. In: *Proceedings of the IEEE/CVF international conference on computer vision*, pp 5229–5238
 36. Tomar D, Lortkipanidze M, Vray G, Bozorgtabar B, Thiran JP (2021) Self-attentive spatial adaptive normalization for cross-modality domain adaptation. *IEEE Trans Med Imaging* 20:20
 37. Wang G, Li W, Ourselin S, Vercauteren T (2017) Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In: *International MICCAI brainlesion workshop*. Springer, pp 178–190
 38. Wang J, Huang H, Chen C, Ma W, Huang Y, Ding X (2019a) Multi-sequence cardiac MR segmentation with adversarial domain adaptation network. In: *International workshop on statistical atlases and computational models of the heart*. Springer, pp 254–262
 39. Wang J, Sun K, Cheng T, Jiang B, Deng C, Zhao Y, Liu D, Mu Y, Tan M, Wang X et al (2020) Deep high-resolution representation learning for visual recognition. *IEEE Trans Pattern Anal Mach Intell* 20:20
 40. Wang X, Yang S, Fang Y, Wei Y, Wang M, Zhang J, Han X (2021) Sk-unet: an improved u-net model with selective kernel for the segmentation of LGE cardiac MR images. *IEEE Sens J* 20:20
 41. Wang Y, Li C, Zhu T, Zhang J (2019) Multimodal brain tumor image segmentation using wrn-ppnet. *Comput Med Imaging Graph* 75:56–65
 42. Wu F, Zhuang X (2020) Cf distance: a new domain discrepancy metric and application to explicit domain adaptation for cross-modality cardiac image segmentation. *IEEE Trans Med Imaging* 39(12):4274–4285
 43. Yang G, Zhuang X, Khan H, Haldar S, Nyktari E, Li L, Wage R, Ye X, Slabaugh G, Mohiaddin R et al (2018) Fully automatic segmentation and objective assessment of atrial scars for long-standing persistent atrial fibrillation patients using late gadolinium-enhanced mri. *Med Phys* 45(4):1562–1576
 44. Yu H, Zha S, Huangfu Y, Chen C, Ding M, Li J (2020) Dual attention u-net for multi-sequence cardiac MR images segmentation. *Myocardial pathology segmentation combining multi-sequence CMR challenge*. Springer, Berlin, pp 118–127
 45. Yuan Y, Chen X, Chen X, Wang J (2021) Segmentation transformer: object-contextual representations for semantic segmentation. In: *European conference on computer vision (ECCV)*, vol 1
 46. Zhai S, Gu R, Lei W, Wang G (2020) Myocardial edema and scar segmentation using a coarse-to-fine framework with weighted ensemble. *Myocardial pathology segmentation combining multi-sequence CMR challenge*. Springer, Berlin, pp 49–59
 47. Zhang D, Huang G, Zhang Q, Han J, Han J, Wang Y, Yu Y (2020) Exploring task structure for brain tumor segmentation from multi-modality MR images. *IEEE Trans Image Process* 29:9032–9043
 48. 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
 49. Zhang J, Xie Y, Liao Z, Verjans J, Xia Y (2020) Efficientseg: a simple but efficient solution to myocardial pathology segmentation challenge. *Myocardial pathology segmentation combining multi-sequence CMR challenge*. Springer, Berlin, pp 17–25
 50. Zhang L, Zhang J, Shen P, Zhu G, Li P, Lu X, Zhang H, Shah SA, Bennamoun M (2020) Block level skip connections across cascaded v-net for multi-organ segmentation. *IEEE Trans Med Imaging* 39(9):2782–2793
 51. Zhang X, Noga M, Punithakumar K (2020) Fully automated deep learning based segmentation of normal, infarcted and edema regions from multiple cardiac MRI sequences. *Myocardial pathology segmentation combining multi-sequence CMR challenge*. Springer, Berlin, pp 82–91
 52. Zhang Y, Wu J, Liu Y, Chen Y, Chen W, Wu EX, Li C, Tang X (2021) A deep learning framework for pancreas segmentation with multi-atlas registration and 3d level-set. *Med Image Anal* 68:101884
 53. Zhang Z, Liu C, Ding W, Wang S, Pei C, Yang M, Huang L (2020) Multi-modality pathology segmentation framework: application to cardiac magnetic resonance images. *Myocardial pathology segmentation combining multi-sequence CMR challenge*. Springer, Berlin, pp 37–48
 54. Zhou T, Ruan S, Canu S (2019) A review: deep learning for medical image segmentation using multi-modality fusion. *Array* 3:100004
 55. Zhuang X (2018) Multivariate mixture model for myocardial segmentation combining multi-source images. *IEEE Trans Pattern Anal Mach Intell* 41(12):2933–2946

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.