Segmenting Medical MRI via Recurrent Decoding Cell

Ying Wen[1], Kai Xie[1], Lianghua He[2*]

[1]East China Normal University, Shanghai, China and [2]Tongji University, Shanghai, China

ywen@cs.ecnu.edu.cn, kxie@stu.ecnu.edu.cn, *helianghua@tongji.edu.cn

Introduction

Fig 1: Some visualization results on the medical segmentation task.

There exits three main challenges for medical image segmentation:

- The importance of hierarchical feature fusion.
- The use of multi-modality information.
- The robustness of networks.

In this paper, we propose an encoder-decoder network for medical MRI segmentation. The main contributions are as follows:

- We propose a new feature fusion unit called Recurrent Decoding Cell (RDC), which leverages the ability of convolutional RNN in memorizing long-term context information. The parameters in RDC are shared in each hierarchical stage, therefore, it is a flexible module and can be added into any encoder-decoder segmentation network to help reduce model size.
- We propose a Convolutional Recurrent Decoding Network (CRDN) based on RDC for segmenting multi-modality medical MRI. CRDN utilizes CNN backbone as the feature encoder and RDC-based decoder to form an end-to-end segmentation network. CRDN effectively increases the segmentation accuracy and shows its robustness in image noise and intensity non-uniformity.

Method

Convolutional Recurrent Decoding Network (CRDN). The proposed CRDN is an end-to-end segmentation pipeline which takes multi-modality images as input and produces per-pixel segmentation inference for each tissue. CRDN consists of two phases: the CNN backbone is utilized as the encoder to extract feature maps for hierarchical feature learning and the proposed recurrent decoding cell (RDC) is designed as the decoder to gradually recover the spatial resolution.

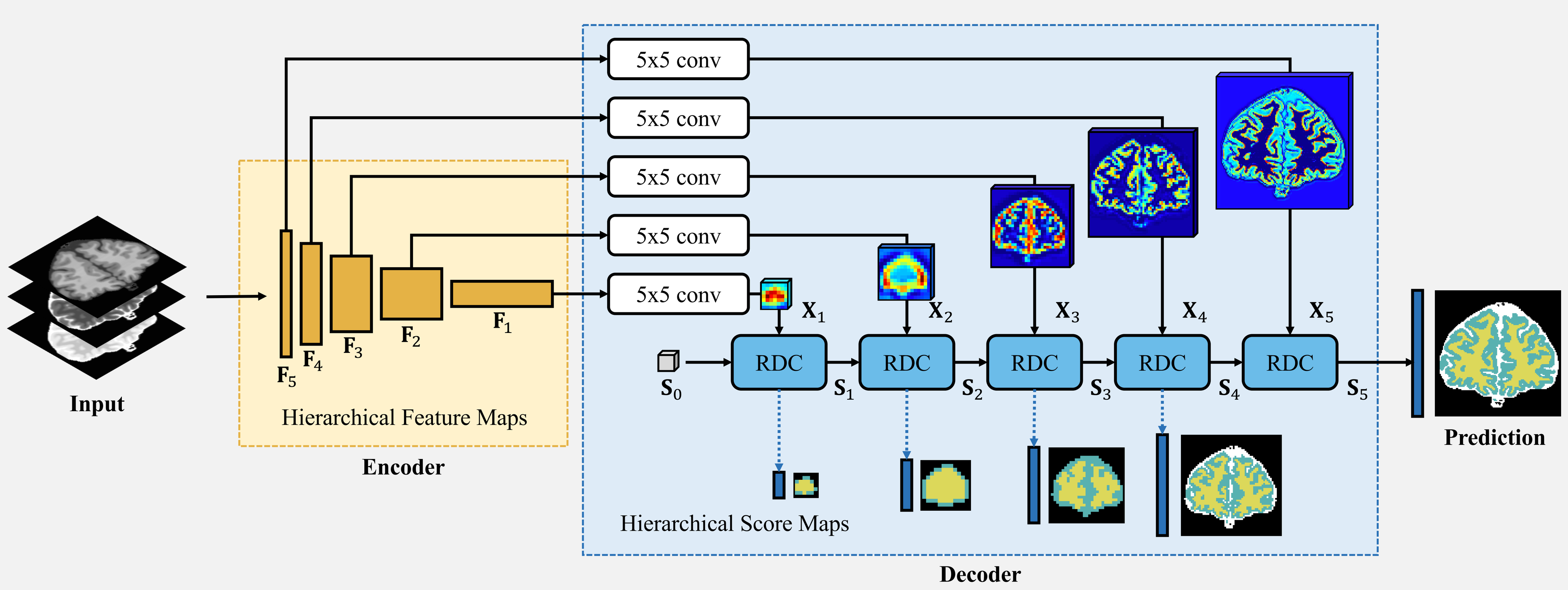
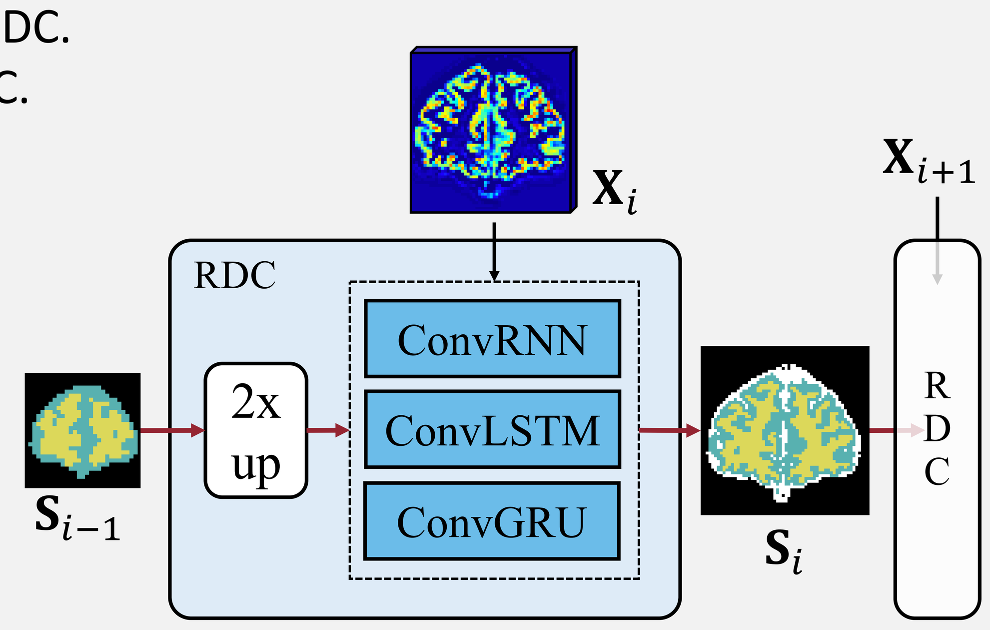


Fig 2: Illustration of the proposed Convolutional Recurrent Decoding Network.

Recurrent Decoding Cell (RDC). In each RDC unit, the previous score map $\mathbf{S}_{i-1}$, which can be treated as the hidden state of a RNN cell, is refined with the current input $\mathbf{X}_i$, generating the current new score map $\mathbf{S}_i$ as the input of the following RDC. Specifically, we first upsample the score map $\mathbf{S}_{i-1}$ to the same spatial dimension as $\mathbf{X}_i$, and this can be done through either bilinear interpolation or learnable transposed convolution. Then, the upsampled score map and the current feature input are fed into a convolutional RNN cell for feature decoding. According to different types of RNNs, three types of RDCs are defined:

- **RDC-ConvRNN.** Basic convolutional RNN is utilized in RDC.
- **RDC-ConvLSTM.** ConvLSTM is utilized in RDC.
- **RDC-ConvGRU.** ConvGRU is utilized in RDC.

Fig 3. The structure of RDC. The previous score map, the current input feature map, the current score map and the next input feature map are denoted as $\mathbf{S}_{i-1}$, $\mathbf{X}_i$, $\mathbf{S}_i$, $\mathbf{X}_{i+1}$, respectively. One of the convolutional RNNs in the dashed box is used for feature fusion.



Results

Datasets. Experiments are implemented on two brain datasets and one cardiovascular MRI dataset: the BrainWeb dataset (Cocosco et al. 1997), the MICCAI 2013 MRBrainS Challenge dataset (Mendrik et al. 2015) and the HVSMR 2016 Challenge dataset (Pace et al. 2015).

Ablation Study. We choose VGG16, ResNet50 and U-Net-like backbones as the encoder and use our three types of RDCs in the decoding phase.

Table 1: Ablation study on BrainWeb, MRBrainS and HVSMR. Evaluated by *Dice Coefficient*.

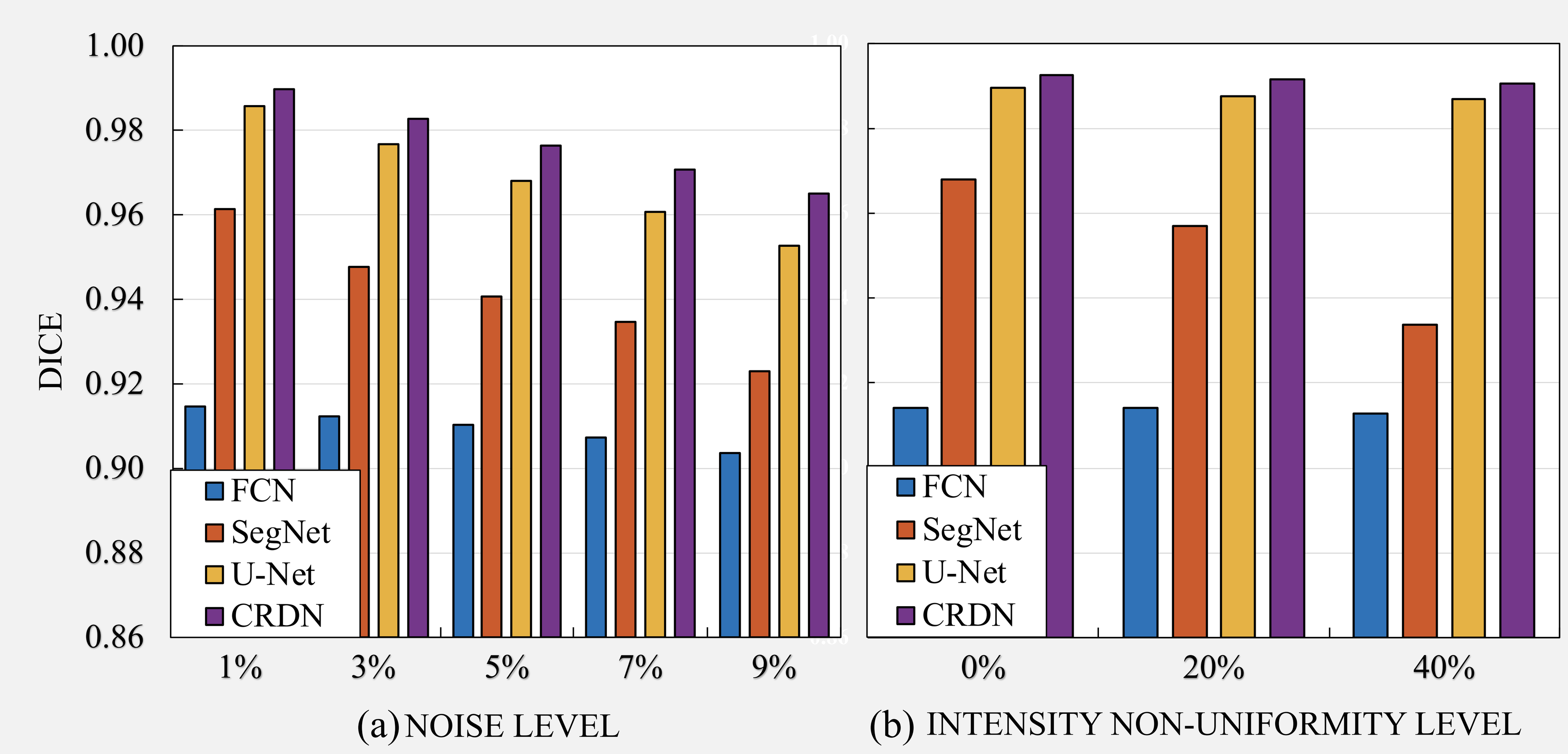
Encoder	Backbone	BrainWeb	MRBrainS	HVSMR
RDC-ConvRNN	VGG16	0.9927	0.9088	0.8813
	ResNet50	0.9920	0.9050	0.8641
	U-Net-Like	0.9934	0.9068	0.8800
RDC-ConvLSTM	VGG16	0.9916	0.9126	0.8641
	ResNet50	0.9896	0.9012	0.8606
	U-Net-Like	0.9919	0.9112	0.8777
RDC-ConvGRU	VGG16	0.9926	0.9061	0.8776
	ResNet50	0.9912	0.9021	0.8696
	U-Net-Like	0.9925	0.9028	0.8796

Comparison with Encoder-decoder Networks. We implement FCN, SegNet, U-Net and our CRDN with VGG16, ResNet50 and U-Net-like backbones as feature encoders.

Table 2: Comparisons on BrainWeb, MRBrainS and HVSMR. Evaluated by *Dice Coefficient*.

Model	BrainWeb	MRBrainS	HVSMR	#Params (240×240×3)
FCN with VGG16	0.9142	0.8637	0.8368	50.42M
SegNet with VGG16	0.9679	0.8294	0.7718	29.42M
U-Net with VGG16	0.9923	0.8991	0.8201	25.86M
CRDN with VGG16	**0.9927**	**0.9126**	**0.8813**	**14.87M**
FCN with ResNet50	0.9115	0.8374	0.8266	115.83M
U-Net with ResNet50	0.9909	0.9039	0.8371	71.86M
CRDN with ResNet50	**0.9920**	**0.9050**	**0.8696**	**23.65M**
FCN with U-Net backbone	0.9176	0.8618	0.8295	**1.19M**
SegNet with U-Net backbone	0.9455	0.8448	0.8099	2.36M
U-Net	0.9892	0.9021	0.8593	1.94M
CRDN with U-Net backbone	**0.9934**	**0.9112**	**0.8800**	1.23M

Network Robustness. (a) Experimental results on **images corrupted with noise**. Note that the noise percentage represents the percent ratio of the standard deviation of the white Gaussian noise versus the signal for a reference tissue. (b) Experimental results on **images with intensity non-uniformity**. Note that for a 20% level, the multiplicative INU field has a range of values of 0.90, ..., 1.10 over the brain area. For other INU levels, the field is linearly scaled accordingly.



Conclusion

The experimental results demonstrate that RDC helps to achieve better boundary adherence compared with other segmentation decoders and reduces the model size. CRDN achieves promising segmentation results for medical image segmentation and shows its robustness to image noise and intensity non-uniformity in MRI.

Poster for AAAI-2020, New York, New York, USA.

3663 N. Zhongshan Rd., Shanghai 20062, China.

PDF CODE