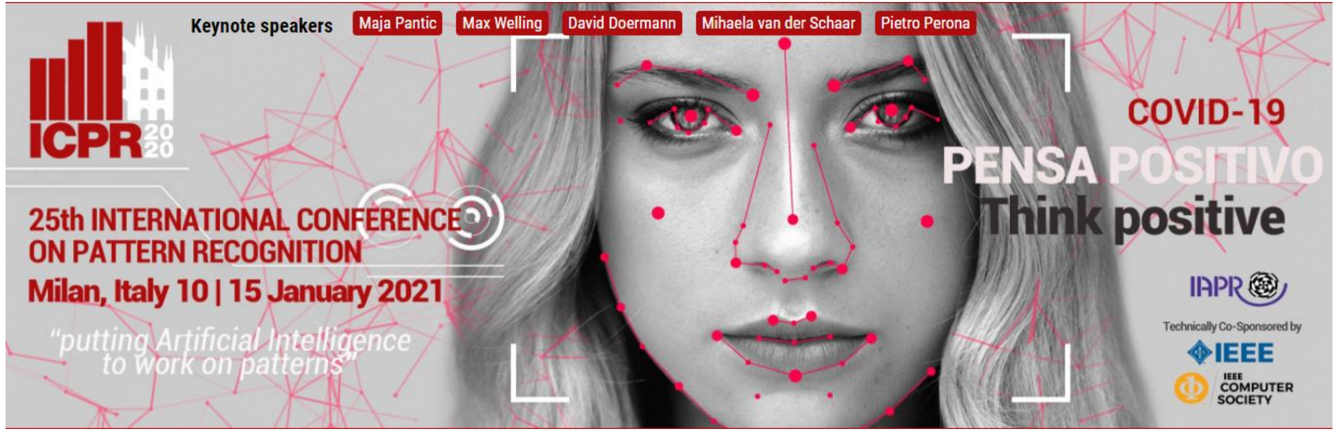

25th International Conference on Pattern Recognition

ICPR2020

BCAU-Net: A Novel Architecture with Binary Channel Attention Module for MRI Brain Segmentation

Yongpei Zhu*[a], Zicong Zhou[b], Guojun Liao[b], Kehong Yuan[a]
[a]Graduate School at Shenzhen, Tsinghua University, Shenzhen, China
[b]The University of Texas at Arlington, Arlington, USA

Jan. 14th,2021

Main Contents

- 1. ABSTRACT
- 2. INTRODUCTION
- 3. METHODS
- 4. RESULTS
- 5. CONCLUSIONS
- 6. REFERENCES

1. ABSTRACT

- **The development of deep learning is slow in magnetic resonance image (MRI) segmentation of normal brain tissues.**
- Contributions of our paper:

(1) We propose a new architecture, Binary Channel Attention U-Net (BCAU-Net), by introducing a novel Binary Channel Attention Module (BCAM) into skip connection of U-Net, which can take full advantages of the channel information extracted from the encoding path and corresponding decoding path.

(2) Spatial pyramid pooling (SPP) modules with different pooling operations are used in BCAM.

(3) We verify this model on two datasets including IBSR and MRBrainS18, and obtain better performance on MRI brain segmentation compared with other methods.

2. INTRODUCTION

- **The precise automatic segmentation of brain tissues such as white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF) of MRI is of great importance for accurate evaluation of early brain development.**
- **Convolutional neural networks (CNNs) have been used for brain tissue segmentation:**
  - **2D U-Net:Ronneberger et al.(Ref. [1]), V-Net:Milletari et al.(Ref. [2]), 3D U-Net:(Ref. [3]), VoxResNet:Chen et al.(Ref. [4])**
- **Attention mechanism has been used in medical image segmentation:**
  - **Binary version of sSE:Roy et al.(Ref. [5]), A novel attention gate(AG):Oktay et al.(Ref. [6]), A connection sensitive attention U-Net (CSAU):Li et al. (Ref. [7])**
- Contributions:
  - We propose a new architecture BCAU-Net by introducing a novel Binary Channel Attention Module (BCAM) to better provide more precise anatomical segmentation.
  - Spatial pyramid pooling (SPP) [8] modules with different pooling operations are used in BCAM to better aggregate multi-scale spatial information of the feature map.

3. METHODS

➢ Attention Module

Squeeze-and-Excitation (SE), Spatial Squeeze and Channel Excitation (cSE), Channel Squeeze and Spatial Excitation (sSE), Mixed-Supervised Dual-Network (MSDN), Convolutional Block Attention Module (CBAM)

➢ Binary Channel Attention Module (BCAM)

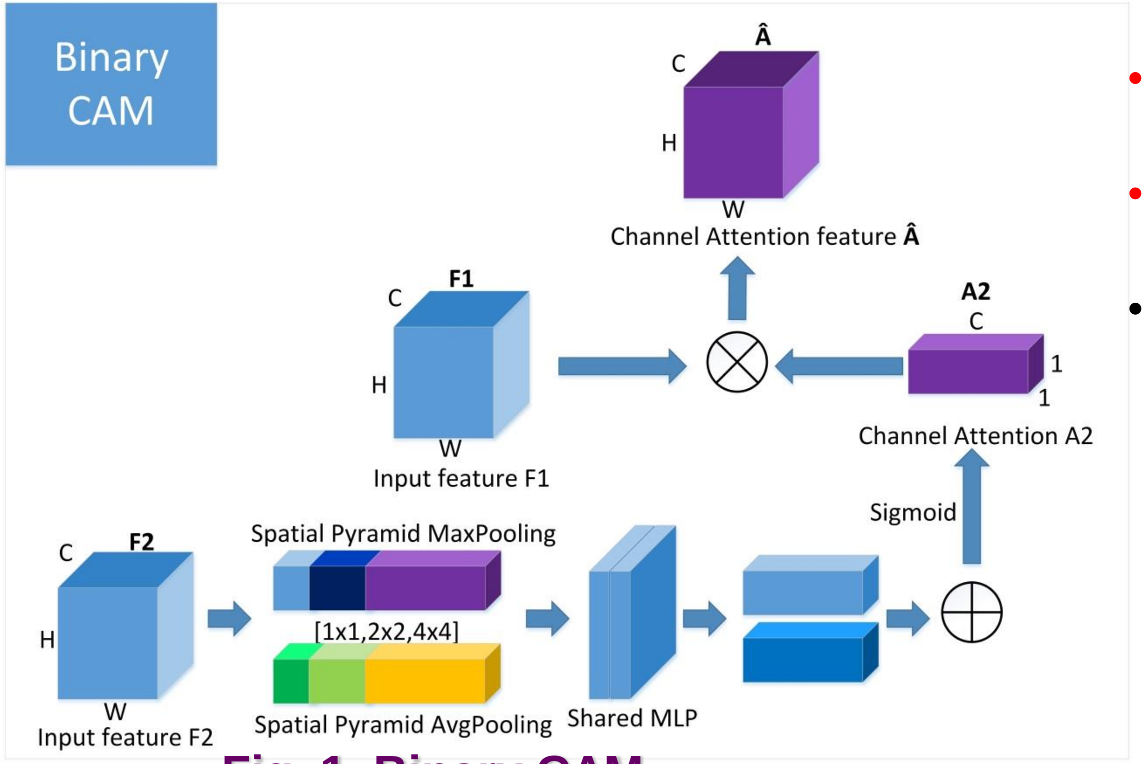

Fig. 1. Binary CAM.

- Spatial Pyramid average-Pooling (SPP with average-pooling operation)
- Spatial Pyramid max-Pooling (SPP with max-pooling operation)
- instead of average-pooling and max-pooling operations: to better aggregate multi-scale spatial information of the feature map (3-level pyramid average-pooling and max-pooling (1x1, 2x2, 4x4) with total 21 bins respectively)

$$
\begin{aligned}
\boldsymbol{A2}(\boldsymbol{F2}) &= \sigma(MLP(SPAvgPooling(\boldsymbol{F2})) \\
&\quad + MLP(SPMaxPooling(\boldsymbol{F2}))) \\
&= \sigma(\boldsymbol{W_1}(\boldsymbol{W_0}(F^c_{spp-avg})) + \boldsymbol{W_1}(\boldsymbol{W_0}(F^c_{spp-max})))
\end{aligned}
$$

$$
\hat{\boldsymbol{A}} = \boldsymbol{A2}(\boldsymbol{F2}) \otimes \boldsymbol{F1}
$$

3. METHODS

➢ Network Structure

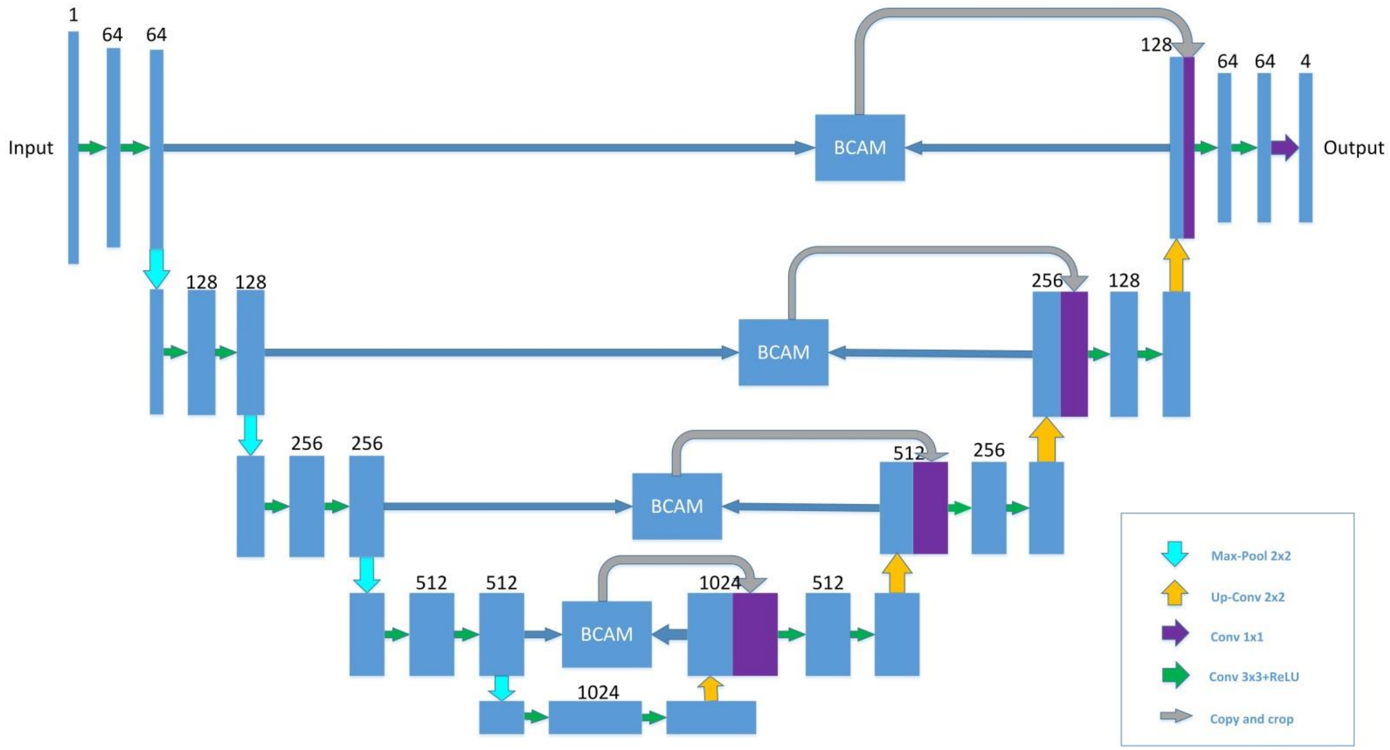

Fig. 2. BCAU-Net architecture.

- The BCAM modules are added into the skip-connection between each encoder stage and corresponding decoder stage.
- The feature map from the output of each BCAM is concatenated with the corresponding feature map that after up-sampling in the decoder stage, utilizing the inter-channel relationship of corresponding low-level feature (low-resolution information after maxpooling) and high-level feature (high-resolution information after up-sampling).

4. RESULTS

➢ Datasets

IBSR:18 MRI volumes (01-18 scans, size:256 × 128 × 256) and the corresponding ground truth(GT)
MRBrainS18: including T1-weighted, T1-IR and T2-FLAIR, seven brain MRI scans (1,4,5,7,14,070,148, size:240 × 240 × 48) with manual segmentations are provided.

➢ Preprocessing

registration, skull stripping, bias field correction, Gaussian smoothed method, intensity normalization and Contrast-Limited Adaptive Histogram Equalization(CLAHE)

➢ Implementation Details

GeForce GTX 1080 Ti GPU、Keras
5-fold cross-validation
loss function： $L = L_{Dice} + \alpha L_{CE}$

$$L_{Dice} = 4 - \sum_{i=0}^{3} \frac{2\sum_{i=1}^{N} p_i y_i + \epsilon}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} y_i + \epsilon}$$

$$L_{CE} = -\frac{1}{N}\left(\sum_{i=1}^{N} p_i y_i + \sum_{i=1}^{N}(1-p_i)(1-y_i)\right)$$

➢ Evaluation Criteria

Dice coefficient [9] (DC) (higher is better), the 95th-percentile of Hausdorff distance (HD) (lower is better) and absolute volume difference (AVD) (lower is better)

$$DC = \frac{2 * (A \cap B)}{|A| + |B|} \qquad H(A,B) = max(h(A,B), h(B,A)) \qquad AVD(A,B) = \frac{\| A - B \|}{\| A \| + \| B \|}$$

4. RESULTS

➢ Experimental Results

TABLE I

RESULTS OF 5-FOLD CROSS VALIDATION ON IBSR DATASET FOR DIFFERENT EXPERIMENTS (DC:%, HD: MM, AVD:%). THE BEST OBTAINED RESULTS ARE PRESENTED IN THE FIVE FOLDS FOR IBSR. BEST RESULTS ARE HIGHEST IN BOLD.

Experiment	DC			Avg. DC	HD			Avg. HD	AVD			Avg. AVD
	CSF	GM	WM		CSF	GM	WM		CSF	GM	WM	
2D U-Net (Baseline) [6]	85.19	90.34	89.20	88.24	3.0835	1.9665	2.4675	2.5058	8.52	1.68	5.21	5.14
3D U-Net [13]	85.51	89.58	89.33	88.48	3.0494	1.7711	1.8949	2.5425	9.90	1.72	5.26	4.56
VoxResNet [4]	71.58	91.37	90.23	84.39	19.1106	1.5159	1.7693	7.4653	18.77	3.84	4.61	9.07
BCAU-Net-R	83.34	91.04	89.77	88.05	3.3411	1.8781	2.2081	2.4758	10.10	1.21	4.00	5.10
BCAU-Net	85.41	91.38	89.78	**88.86**	2.9370	1.8744	2.0858	**2.2991**	7.72	0.97	2.83	**3.84**
BCAU-Net-E	84.18	91.02	89.69	88.30	3.3655	1.8715	2.1059	2.4476	9.43	1.37	4.38	5.06

TABLE II

RESULTS OF 5-FOLD CROSS VALIDATION ON MRBRAINS18 DATASET FOR DIFFERENT EXPERIMENTS (DC:%, HD: MM, AVD:%). THE BEST OBTAINED AVERAGE RESULTS ARE PRESENTED IN THE FIVE FOLDS FOR MRBRAINS18. BEST RESULTS ARE HIGHEST IN BOLD.

Experiment	Avg. DC	Avg. HD	Avg. AVD
2D U-Net (Baseline) [6]	83.93	8.3047	9.05
3D U-Net [13]	83.62	8.3441	9.71
VoxResNet [4]	84.59	7.4064	7.99
BCAU-Net-R	85.32	7.1375	4.66
BCAU-Net	85.73	6.6850	4.15
BCAU-Net-E	**85.75**	**6.6825**	**4.09**

- RCAU-Net produces better results (DC: 88.86%, HD: 2.2991mm, AVD: 3.84%) on IBSR dataset than 2D U-Net (Baseline), with a relative improvement of 0.62% on DC, which shows the effectiveness of BCAM block
- BCAU-Net outperforms other architectures in terms of average DC, HD and AVD on two datasets.
- it help utilize the inter-channel relationship of corresponding low-level and high-level information to better provide more precise anatomical segmentation.

4. RESULTS

➢ Experimental Results

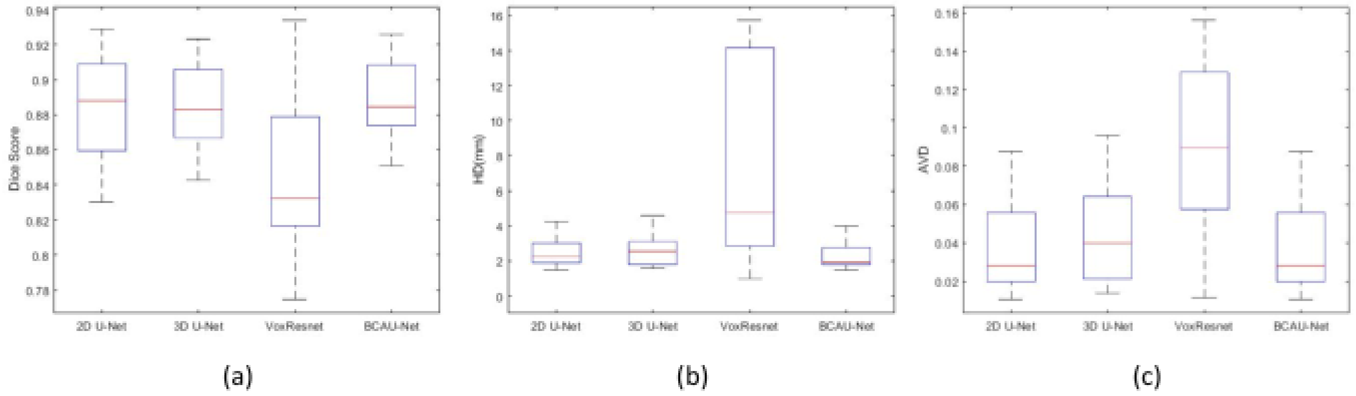

Fig. 3. Boxplots of best obtained results for different models on IBSR: (a) Boxplot of DC, (b) Boxplot of HD, (c) Boxplot of AVD.

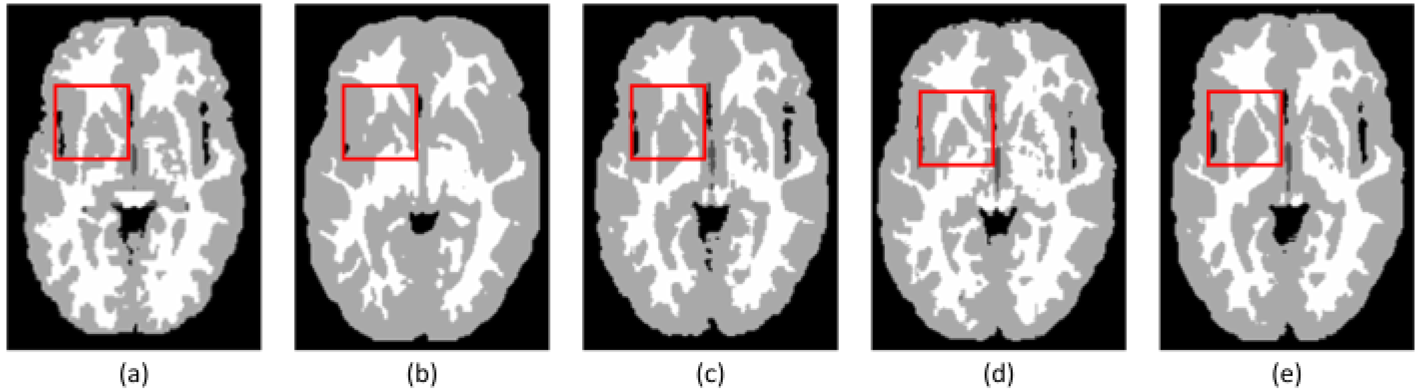

Fig. 4. Best predicted results for different models on IBSR: (a) Label (Ground Truth), (b) 2D U-Net model, (c) 3D U-Net model, (d) VoxResNet model, (e) Our model (BCAU-Net).

- BCAU-Net has the highest average DC, the lowest maximum value of HD and AVD. However, VoxResNet performs the worst among all the methods according to average DC, HD and AVD.
- BCAU-Net get the best segmentation result than other models compared with the Ground Truth which are noted in the red box, and our method shows better structures and details of brain tissue, such as WM.

5. CONCLUSIONS

- Propose a new architecture RCAU-Net by introducing a novel Binary Channel Attention Module (BCAM) into skip connection of U-Net, which can take full advantages of the channel information extracted from the encoding path and corresponding decoding path, it can enhance the segmentation performance by focus on the details and textures of image structures.

- To better aggregate multi-scale spatial information of the feature map, spatial pyramid pooling (SPP) with three pooling windows (1x1,2x2,4x4) are used in BCAM instead of original average-pooling and max-pooling operations.

- We verify this model on two datasets including IBSR and MRBrainS18, and obtain better performance on MRI brain segmentation compared with other methods.

6. REFERENCES

- [1] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in Medical Image Computing and Computer-Assisted Intervention, 2015, pp. 234–241.
- [2] F. Milletari, N. Navab, and S. A. Ahmadi, V-Net: fully convolutional neural networks for volumetric medical image segmentation. In Proceedings of 3DV, 2016, pp. 565–571.
- [3] O. C¸ic¸ek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, ”3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation,” in Medical Image Computing and Computer-Assisted Intervention, 2016, pp. 424–432.
- [4] H. Chen, Q. Dou, L. Yu, J. Qin, and P. A. Heng, “VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images,” NeuroImage, vol. 170, pp. 446–455, 2018.
- [5] A. G. Roy, S. Siddiqui, S. P ¨ olsterl, N. Navab, and C. Wachinger, ‘Squeeze and Excite’ Guided Few-Shot Segmentation of Volumetric Images, unpublished.
- [6] O. Oktay, et al., Attention U-Net: Learning Where to Look for the Pancreas. In MIDL, 2018.
- [7] R. R. Li, M. M. Li, J. C. Li, and Y. T. Zhou, Connection Sensitive Attention U-NET for Accurate Retinal Vessel Segmentation. unpublished.
- [8] K. He, X. Zhang, S. Ren, and J. Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. In ECCV, 2014.
- [9] A. A. Taha and A. Hanbury, Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. Bmc Medical Imaging, vol.15, no.1, pp. 29, 2015.

Thank you for Listening!