AW-Net: A Novel Fully Connected Attention-based Medical Image Segmentation Model

Debojyoti Pal, Tanushree Meena, Dwarikanath Mahapatra, Sudipta Roy
Jio Institute in Navi Mumbai, Maharashtra, India

ICCV23 PARIS

INTRODUCTION

- ❑ **Multimodal medical imaging** poses a unique challenge to the data scientist looking at that data, since it is not only voluminous, but also extremely **heterogenous**. As a result, the task of detecting and segmenting intricate anomalies such as **tumors** and **target lesions** proves to be difficult.
- ❑ We aim to introduce a novel, **generalized**, attention-based segmentation model, namely **AW-Net** to accomplish the task of **multiorgan segmentation** from **multimodal** medical images.

The main contributions of this paper are as follows:

1. Novel regularized transient block **(RTB)**, comprising of regularized convolutional blocks and **dropout layers**.
2. Experiments performed on **multimodal, benchmark** datasets: **RSNA, BraTS, DUKE, and QIN**.
3. A study of the **computational cost** of the model in an attempt to showcase the **cost-effectiveness** of the model.

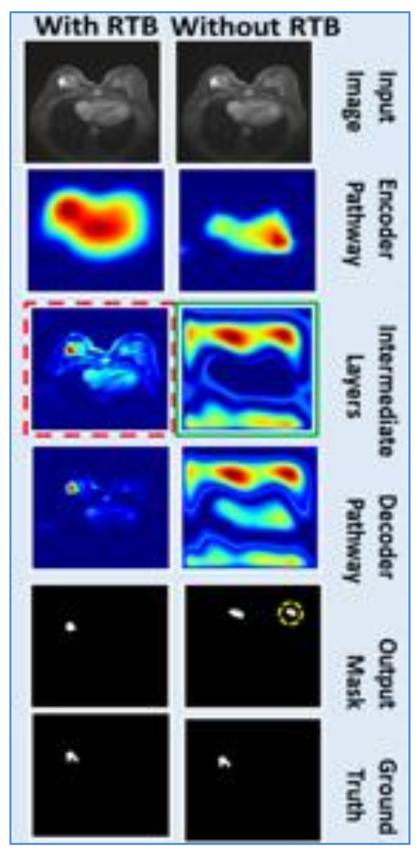


Fig. 1. Feature maps showing the effectiveness of RTB

DATA-GROWTH STUDY

- **The primary objective of this study is to analyze how the model performs when trained on a reduced dataset and evaluating on the remaining data.** These conditions, specifically explores the impact of data size on its effectiveness.
- The study aims to uncover valuable insights about the direct relationship between dataset size and model performance.
- The investigation commences by utilizing 30% of the data for training and reserving 70% for testing. Through a systematic approach, the training data is incrementally increased by 5%, while the testing data is simultaneously reduced by 5%, until reaching a configuration with 80% of the data for training and 20% for test.

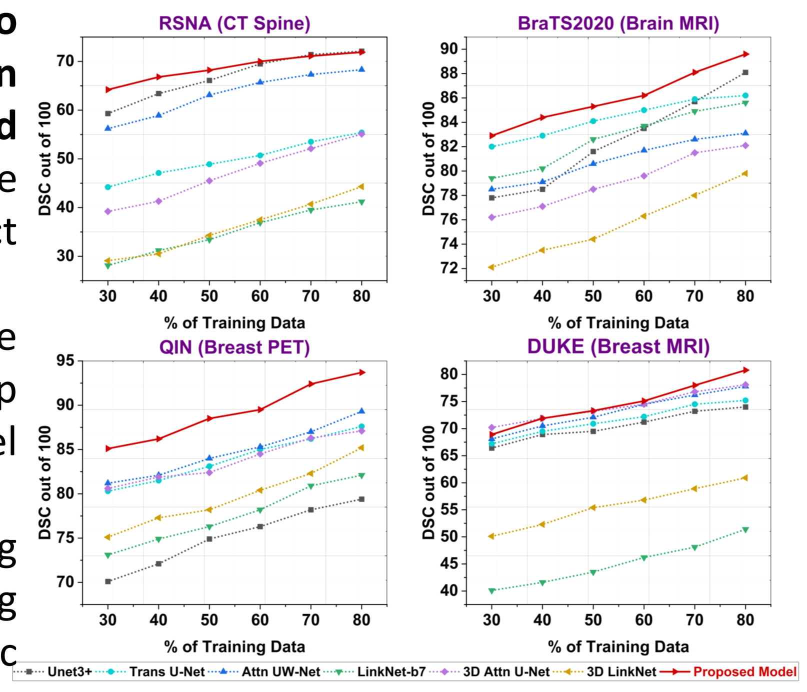
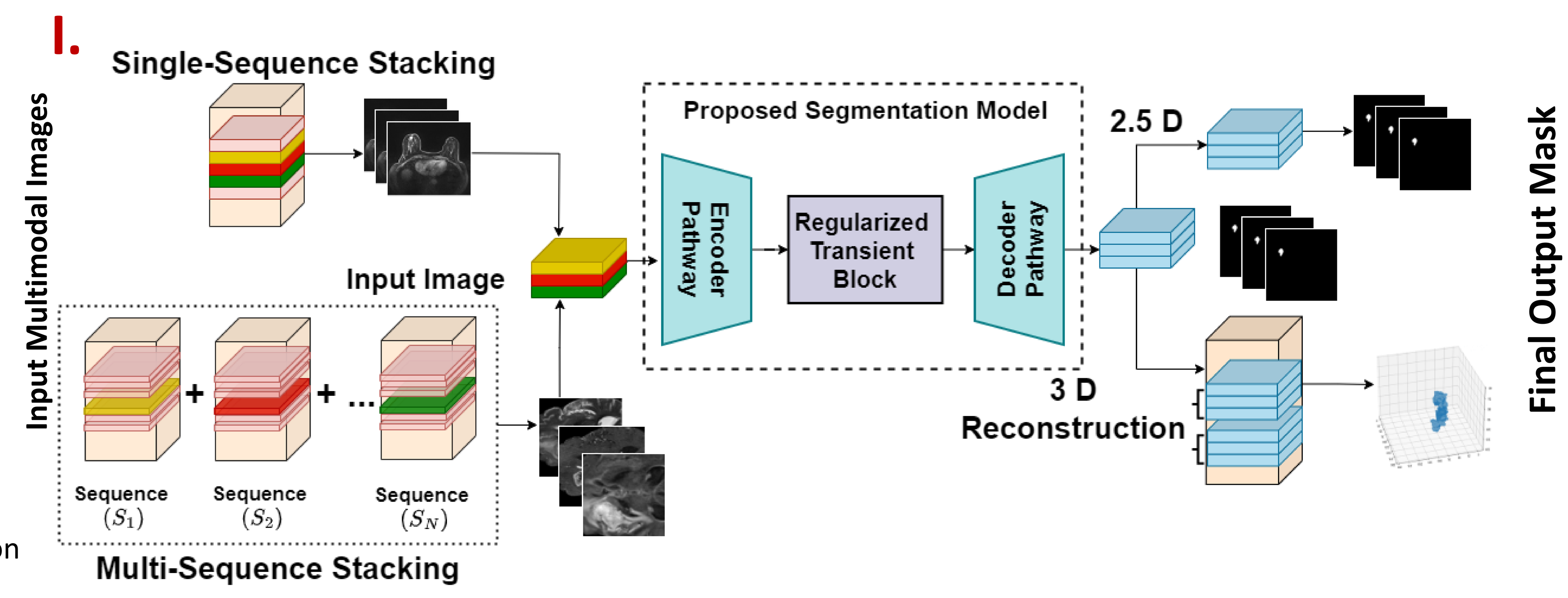


Figure 4: Line plots illustrating the performance of AW-Net for the data growth study with respect to different datasets. A) RSNA. B) BraTS2020. C) QIN. D) DUKE.

METHODOLOGY

❑ **The methodology involved in the multimodal medical segmentation comprises of following parts:**

1. **2.5D Stacking:** This involves stacking of multiple sequences **(multi-sequence stacking)** or consecutive images of a single modality **(single-sequence stacking)** based on the given input data.
2. ***Proposed Model***: The proposed model comprises of **encoding pathway, decoding pathway,** and the novel regularized transient block **(RTB)** which acts as a bridge between these two pathways.
3. The output is then stacked together to provide a **reconstructed 3D** version of the target lesion/ tumor.
4. ***Encoding Pathway*** comprises of four fully connected encoder block connected by the max-pooling layer. The first three blocks comprises of two convolution layers with a 3×3 filter size stacked together. Each convolution layers are followed by a ReLu activation and a Batch Normalization operation. These convolution layers facilitate the extraction of important features in subsequent blocks. The feature map of the convolution layers increases by a factor of 2 for each subsequent layer from 32 to 256.
5. ***Novel Regularized Transient Block:*** One of the major problems of attention U-Net is the bottleneck. Bottleneck refers to the narrowest part of the network, where the spatial resolution of the feature maps is reduced. The reduced spatial resolution leads to loss of fine-grained details important for accurate segmentation. The bottleneck of the U-Net architecture is characterized by an increased number of channels, resulting in feature-rich vectors. However, this can lead to two issues: first, an elevated risk of overfitting, and second, an increase in the computational complexity of the model. To solve these problems, we propose a fully connected layer, named regularized transient block (RTB) as a replacement of the bottleneck.
6. ***Decoder Pathway*** consists of four fully connected decoder blocks connected via an up-sampling layer. Each decoder block has an up-sampled vector and an attention vector as an input. The attention vector is basically an output of the attention gate, whereas the up-sampled vector is the output of a transverse convolution layer of filter size 2×2. Both these vectors are concatenated and subsequently serve as an input vector for a series of two convolution layer stacked on top of each other.



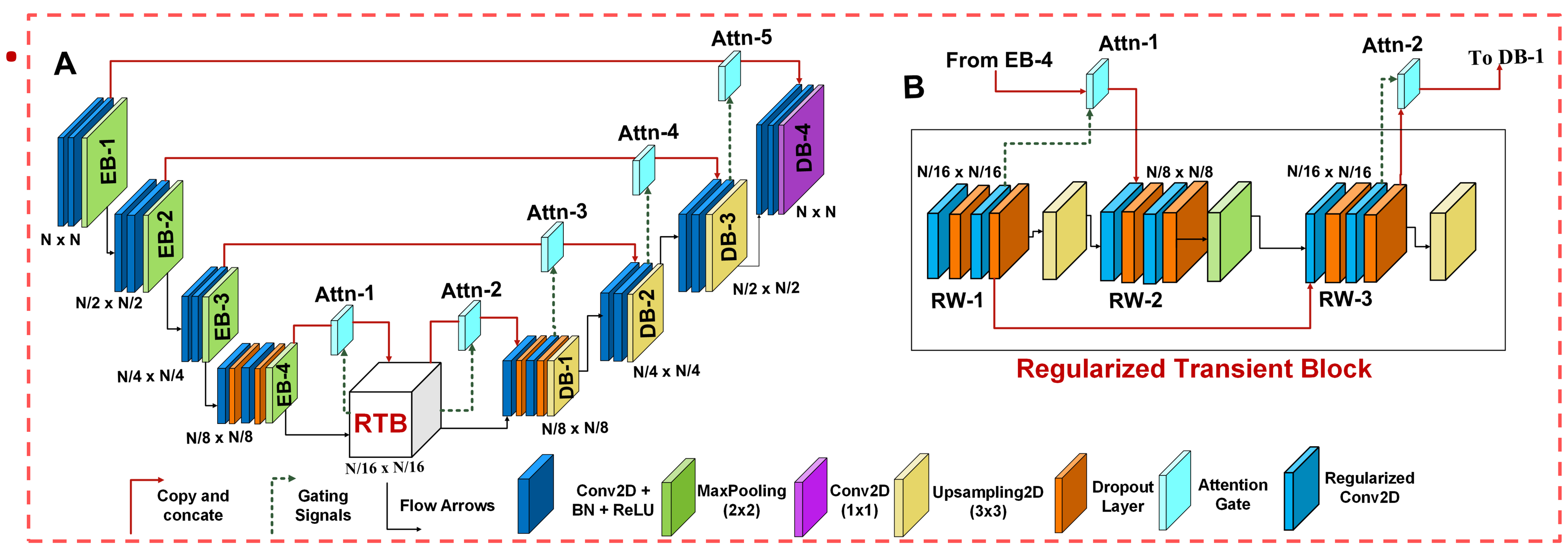


Figure 2 (I) The workflow of the segmentation process showcasing the 2.5 Stacking in green and the proposed architecture in red (II) The detailed architecture of the AW-Net. A) Block diagram illustrating the AW-Net. B) Layer-wise representation of the RTB.

RESULTS

Figure 3 represents the qualitative analysis of the generated predicted mask for different modalities such as MRI (breast and brain), PET (breast), and CT scan (Cervical Spine) from top to bottom. To prove the efficiency and the robustness of the model, Dice Similarity Coefficient (DSC), Average Hausdorff distance (HD), and False Positive Rate (FPR) are used as a performance metrics. The proposed AW-Net achieves an average DSC of more than 80. The maximum (Max), minimum (Min), average (Avg) and standard deviation (SD) values in terms of DSC, HD and FPR for the different modalities and sequences is shown in Table1.

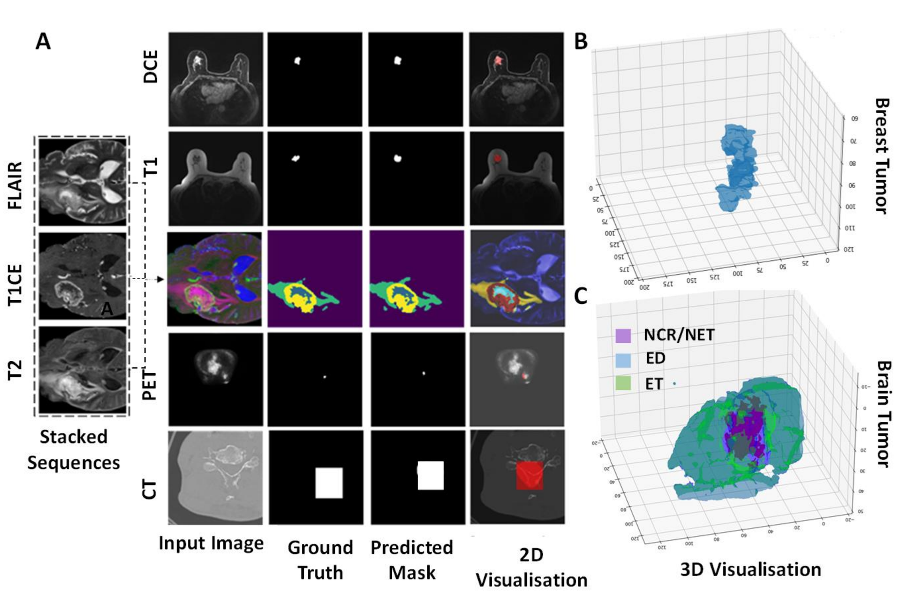


Fig. 3. A detailed analysis of the predicted mask generated by the proposed AW-Net with respect to A) 2D visualization of individual segmented masks for a single image slice, B) 3D visualization of predicted breast tumor and C) 3D visualization of brain tumor.

An average DSC of 81.3, 80.5 and 89.6 is reported for DCE-MRI, T1-MRI, and TICE+T2+FLAIR MRI, respectively. For other modalities such as Breast PET and Cervical Spine CT an average DSC of 93.7 and 71.9 are reported, respectively. An average HD of 10.7 and 9.9 is observed for DCE and T1 for breast MRI and 10.5 in brain MRI. Similarly, an average HD of 0.5 and 6.2 is observed for breast PET and cervical spine CT. Table 1 reports an average FPR of 0.04 for breast MRI segmentation, 0.1, 0.01 and 3.5 for brain MRI and breast PET and cervical spine CT, respectively.

Table 1 Summarized results obtained by the proposed AW-Net.

Body Parts Modality	Sequences	DSC (%)	HD (mm)	FPR (mm)
Breast MRI	DCE	81.3 ± 12.3	10.7 ± 0.7	0.04 ± 0.01
		89.6 - 61.5	11.3 - 9.7	0.15 - 0.01
	T1	80.5 ± 11.5	9.9 ± 0.6	0.04 ± 0.02
		87.5 - 50.2	11.8 - 9.1	0.16 - 0.01
Brain MRI	T1CE+ T2+FLAIR	89.6 ± 8.6	10.5 ± 0.5	0.1 ± 0.08
		96.8 - 59.5	11.3 - 8.7	0.18 - 0.08
Breast PET	-	93.7 ± 11.0	00.5 ± 0.6	0.01 ± 0.05
		96.8 - 60.3	02.0 - 0.5	0.14 - 0.02
Cervical Spine CT	-	71.9 ± 22.1	6.2 ± 1.1	3.5 ± 1.2
		95.0 - 25.6	9.9 - 2.2	7.2 - 2.1

Ablation Study and Computational Time

To evaluate the efficiency and effectiveness of AW-Net, studies are performed based on the memory involved while training a single epoch (in Mb) with respect to the number of floating-point operations performed per second (FLOPS) have also been calculated for these segmentation models to induce the implementation of the developed model for real-time applications. The results for the proposed AW-Net with respect to SOTA segmentation models is reported in Figure 4 Training AW-Net for a single epoch takes 62.5 Mb and involves 5.7 G of FLOPS which is the least among other SOTA models. The proposed model also has the least test inference time of 0.13 seconds, making it suitable for real-world applications. This study establishes the fact that RTB with the combination of L1 regularizer and dropout plays a crucial role in improving model performance and reducing FPR.

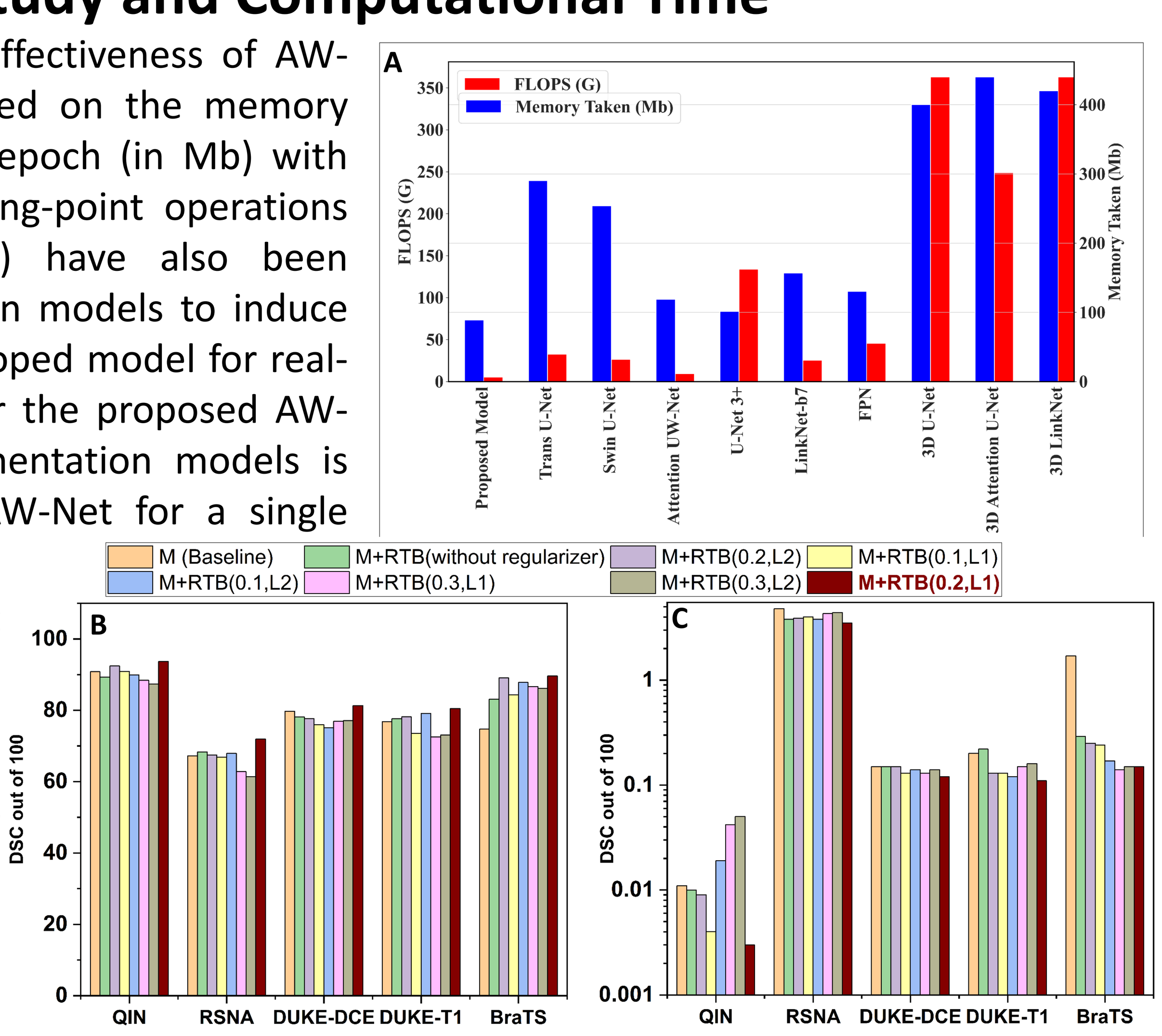


Figure 4: A) Complexity Analysis of the proposed AW-Net. Stacked plot representing B) Average DSC (in %) and C) FPR (in %) of the different models in the ablation study on the proposed AW-Net for different modalities.

SUMMARY AND CONCLUSIONS

- ❑ In this work, we have proposed a novel AW-Net for the segmentation of multimodal 3D/4D images. The AW-Net considers the anatomical features and reduces pixel misclassification by the introduction of RTB. The regularized convolutional layers in the RTB not only reduces the computational complexity but also makes the model robust.
- ❑ Experiments performed on multiple dataset having different modalities and data sequences demonstrate the effectiveness and the generalizability of the model. The model consistently outperforms other benchmark segmentation models in terms of DSC and FPR for public benchmark datasets.

REFERENCES

1. Debojyoti Pal, Pailla Balakrishna Reddy, and Sudipta Roy. Attention uw-net: A fully connected model for automatic segmentation and annotation of chest x-ray. Computers in Biology and Medicine, 150:106083, 2022.
2. Lei Bi, Jinman Kim, Ashnil Kumar, Michael Fulham, and Dagan Feng. Stacked fully convolutional networks with multi-channel learning: application to medical image segmentation. The Visual Computer, 33:1061–1071, 2017.
3. Huahong Zhang, Alessandra M Valcarcel, Rohit Bakshi, Renxin Chu, Francesca Bagnato, Russell T Shinohara, Kilian Hett, and Ipek Oguz. Multiple sclerosis lesion segmentation with tiramisu and 2.5 d stacked slices. In Medical Image Computing and Computer Assisted Intervention– MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part III 22, pages 338–346. Springer, 2019. 2