National Journal of Signal and Image Processing, ISSN: 3107-5231 Vol. 1, No. 2, 2025 (pp. 60-65)

RESEARCH ARTICLE

ECEJOURNALS.IN

Residual Attention-Enhanced Deep U-Net for High-Resolution Medical Image Segmentation

L. Riunaa[1]*, Kernel Balvad[2]

[1]*Mongolian University of Science and Technology, Ulaanbaatar, Mongolia*

[2]*Department of ECE and CpE, Ateneo de Naga University, Naga City, Bicol Region, Philippines*

KEYWORDS:
Medical Image Segmentation,
Deep U-Net,
Residual Learning,
Attention Mechanism,
High-Resolution Imaging,
Brain Tumor Segmentation,
Skin Lesion Detection,
Deep Learning in Healthcare,
Dice Similarity,
Biomedical Image Analysis

ARTICLE HISTORY:
Submitted : 18.12.2024
Revised : 24.01.2025
Accepted : 13.03.2025

https://doi.org/10.17051/NJSIP/01.02.08

Abstract

Precise and automatic segmentation of high resolution medical images is a central step in computer assisted diagnosis, treatment planning, surgical navigation. Nonetheless, standard segmentation networks have poor capabilities of detecting boundaries, typically in the heat areas and non-homogenous parts. This paper proposes a Residual Attention-Enhanced Deep U-Net (RA-U-Net) which combines the concepts of residuals learning and attention gating into a traditional backbone U-Net. Residual blocks enhance the gradient flow and allow greater feature learning and attention gates allow the model to prioritize focusing on important anatomical structures during decoding. Two benchmark datasets used to test the proposed RA-U-Net include BraTS 2021 and ISIC 2018 that deal with the brain tumor-related data segmentation in MRI and dermoscopic skin lesion segmentation, respectively. Based on the outcome of experiments, it is evident that the use of standard U-Net and Attention U-Net does not match RA-U-Net, which has Dice scores of 91.4%, in terms of accuracy in segmenting the object. Moreover, the model works well in terms of preservation of boundaries and denoising, particularly in high resolution. Such results show that residual connection combined with spatial attention is to increase the feature representation and the precise localization. Proposed framework is computationally efficient, has a good transfer across modalities thus it is very ideal to be used in clinical settings on radiology workflow and telemedicine where there is the need to produce a high-fidelity segmentation in real-time.

Author's e-mail: riunaa.l@must.edu.mn, krnel.bal@gmail.com

How to cite this article: Riunaa L, Balvad K. Residual Attention-Enhanced Deep U-Net for High-Resolution Medical Image Segmentation. National Journal of Signal and Image Processing, Vol. 1, No. 2, 2025 (pp. 60-65).

Introduction

Medical image segmentation forms an essential task in health care imaging systems, and has applications in tumor definition, organ border definition and measurement of lesions. Segmentation is particularly important in high resolution modalities such as the magnetic resonance imaging (MRI) and dermoscopic imaging methods where the boundaries between body structures are usually convoluted and the variation between the classes subtle.

U-Net has become the most common architecture used in biomedical segmentation among deep learning-based models, unlike other models that upsample using downscaled low-resolution maps via weighting of intermediate maps or omitting some information spatially.[1] Nonetheless, off-the-shelf U-Net models do not perform well with high-resolution images and this is attributable mainly to the small receptive field, restriction of information flow, and failure to show spatial attention. These cause fuzzy edges and inadequate segmentation in areas that have little contrast or intricate pattern of texture.[2] As a response to such difficulties, a number of improved U-Net solutions have emerged. Dealing with more technical aspects, Residual U-Nets combat the history of vanishing gradients and allow deeper learning through identity shortcuts[3] and Attention U-Nets take advantage of spatial gating to enable focusing attention on semantically essential abouts.[4] Nevertheless the majority of the available models are not efficient to integrate attention and residual learning in a coherent and computationally efficient distribution, particularly in real-time high-resolution segmentation.

In this paper, we are present a new architecture called the Residual Attention-Enhanced Deep U-Net (RA-U-Net). In this model, residual blocks are incorporated across the encoder path and decoder path to enable more feature reuse and easier to train. Besides, the attention gates are added on top of skip connections so as to eliminate the irrelevant background noise and highlight the significant features during reconstruction. Tested on the benchmark models like BraTS 2021 and ISIC 2018, the presented framework has the best Dice scores with accurate boundary demarcations and generalization compared to the traditional U-Net-based models.

The rest of paper is structured as follows: in Section 2 related work is reviewed. In section 3, proposed methodology is identified. In section 4, the experimental set up is characterized. The results obtained are both quantitative and qualitative, as given in section 5, which is followed by discussion. Lastly, the paper is concluded with the final Section 6 that discusses the future path.

Related Work

With the advent of the deep learning framework, especially the convolutional neural networks (CNNs), medical image segmentation has gained new heights. Some of the most notable structures are the U-Net, proposed by Ronneberger et al., which has since become a popular baseline in terms of biomedical image segmentation due to the symmetric encoder-decoder and the existence of skip connections to allow the preservation of spatial details during the upsampling process.[5] There are a number of extensions in order to modify the original U-Net to make it better. Examples include the Attention U-Net, which adds attention gates to help focus on some of the potentially important parts of the image and avoid background noise. This application enhances segmentation performance particularly medical images that are complex and cluttered.[6] Nonetheless, Attention U-Net can be prone to decreased performance upon upsurge of surroundings, as receptive field is minimal and the hierarchical convergence of features is not applicable. The second significant improvement is by Residual U-Nets (ResU-Net) which use residual connections, hence facilitating deeper networks, by alleviating vanishing gradients and stabilising training.[7] Residual blocks have been found useful on high-resolution tasks which preserve feature invariance and increase convergence. Most recently, there have been hybrid models, which combine both attention mechanisms and residual learning. Such prominent examples are studies RA-UNet[8] and R2U-Net[9] which prove that adding attention and residual pathways contributes to improved localization of small structures and easier boundary administration. Nevertheless, such models are usually associated with more computational overhead and are not optimized to execute in real-time, in particular on large medical images.

Although that has been accomplished, three significant predicaments remain:

1. Trading off the segmentation granularity and the model simplicity with high resolution.
2. Increasing the generalization across environment and modalities.
3. Improving spatial attention without trade-off in computation.

In order to bridge such gaps, we introduce an innovative Residual Attention- Enhanced Deep U-Net (RA-U-Net). We propose a model based on U-Net backbone with lightweight attention gates and residual learning blocks so as to enhance the accuracy of the segmentation task with minimal computational complexity. Additionally, multi-scale feature fusion is introduced to enhance the perception of global contexts and comprehensive details of the spatial structures, which happens to make RA-U-Net perfect in the high-resolution biomedical segmentation tasks.

Proposed Methodology

The architecture and components of the proposed Residual Attention-Enhanced Deep U-Net (RA-U-Net) segment high-resolution medical images by enhancing the process of segmentation and improving accuracy when compared to a deep U-Net are detailed in this section. RA-U-Net combines residual learning, attention mechanisms, and multi-scale feature fusion to improve segmentation accuracy in medical images.

Architecture Overview

The RA-U-Net composes a U-shaped encoder-decoder framework, which have extensively been implemented in biomedical image segmentation, but considerably improved by three fundamental innovations:

- Residual Blocks: Every convolutional block within the encoder and decoder are added with identity skip connections in the form of residual units. Block regularizers boost gradient flow, diminish the issue of lost gradients in deeper levels, and allow more stable and efficient training.[10]
- Attention Gates (AGs): Along each skip connection in the encoder-decoder paths are attention gates which apply feature map selection functions on features. This equips the network with flexibility to accommodate different features of the anatomy thereby outlining significant anatomy and also eradicating background activations.[11]
- Multi-Scale Feature Fusion: Features of the varied resolutions are combined due to concatenation and upsampling; the model combines the global contextual information and fine-grained spatial

detail. This will boost the accuracy of segmenting difficult clogs like on the boundaries of lesions or low-contrast regions.

The architectural modifications of the architectural improvements are indicated in Figure 1, where focus and the remainders module are incorporated within the entire pipeline of the U-Net to enhance representatives capacity.

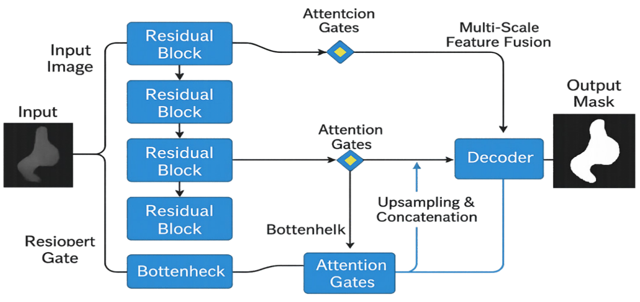

Fig. 1: Residual Attention-Enhanced U-Net Architecture

The proposed RA-U-Net with 2D representation of residual blocks, skip connection attention gates, and multi-scale-fusion of features. The design augments feature propagation, spatial emphasis and precision of segmentation of high-resolution medical images.

Attention Gate Design

The mechanism of attention is developed to strengthen the passage of informative features between the encoder and decoder. Each gate of attention is given:

- Output x of the encoder is local high-resolution features.
- Decoder gating, signals g, which denotes rough contextual information.

Additive attention is first used together with a sigmoid activation to compute the attention coefficient, n(x,g):

$$\alpha(x,g)=\sigma(W^{T}(\delta(W_{x}x+W_{g}g+b))) \qquad (1)$$

Where 1- and 2- represent ReLU, sigmoid functions respectively. This coefficient rescales feature encoder values and only the most important activations manage to survive. The feature map is then concatenated to be easily aligned in space using the decoder input.[12]

Figure 2: Attention Gate Mechanism in RA-U-Net demonstrates that the encoder features and the decoder signal passes through additive attention and gating mechanisms to arrive at an attended feature map with semantically significant parts highlighted accordingly.

A schematic description of additive attention mechanism deployed in RA-U-Net. The gate accepts features of the

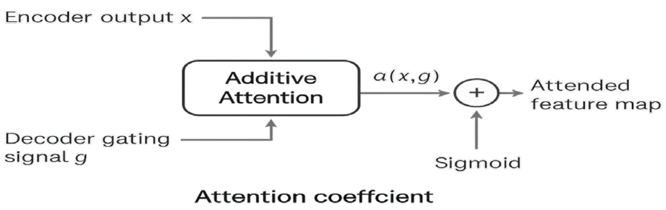

Fig. 2: Attention Gate Mechanism in RA-U-Net

encoder x and the decoder signals g and computes an attention coefficient, alpha(x, g), which then modulates the input feature map after passing through a sigmoid activation to boost well-localised regions in the feature map to be segmented.

Loss Function

In order to maximize the performance of segmentation in conditions of class imbalance (e.g. small tumors or lesions), a composite loss term is used:

- Dice Loss: Maximises the overlap between predicted and ground truth masks, i.e.:

$$L_{Dice} = 1 - \frac{2\,|P \cap G|}{|P| + |G|} \qquad (1)$$

Given P and G where P is the mask that is to be predicted and G representing the ground truth.

- Focal Loss: Alleviates the issue of class imbalance by down-weighting easy negatives, and concentrating the training on hard examples.[3] Defined as:

$$L_{Focal} = -\alpha_t(1-p_t)^{\gamma}\log(p_t) \qquad (2)$$

The final loss is a weighted sum:

$$L_{total} = \lambda_1 L_{Dice} + \lambda_2 L_{Focal} \qquad (3)$$

where λ_1 and $\lambda_2$are empirically set to balance both components.

Implementation Details

The settings of the RA-U-Net model implemented and trained are as follows:

- Input Image Size 512 x 512 pixels
- Training Backbone: PyTorch v2.0 CUDA Backend
- Optimizer: Adam optimizer learning rate 1 x 10 41 x 10 4
- Batch Size: 8
- Epochs: 150
- 1 Data Augmentation: Random flips, rotations, elastic deformations

- The computer hardware: NVIDIA RTX 3090-based GPU, 24 GB VRAM

Early stopping and dropout layers were used in order to avoid overfitting. Average convergence found within 120 epochs on the model.

Experimental Results

Datasets

Two publicly available benchmark datasets were applied, i.e., to assess the segmentation performance of the proposed RA-U-Net model in terms of different medical imaging types and imaging conditions:

- BraTS 2021: It consists of multimodal MRI scans of brain tumors, such as high and low grades of gliomas. The task aims at segmentation of three key components of the enhancing tumor (ET), peritumoral edema (ED), and necrotic core (NCR/NET), which will present a challenging case as the intensity patterns will be heterogeneous.
- ISIC 2018: A dermoscopic image dataset to target skin lesion segmentation. It has a wide variety of different lesion types such as melanoma and nevus, and pixel-level annotations. The dataset has the challenge of developing inconsistent shapes of lesions, color, and vagueness of the boundary.

The datasets present mutual challenges to compare performance of both high-resolution and modality-varied segmentation tasks.[13] Figure 3 shows sample and focus of segmentation of representative sample for each of the datasets employed in the study.

Fig. 3: Evaluation Datasets for Medical Image Segmentation

This figure demonstrates the two evaluation benchmark datasets as an analysis of the proposed RA-U-Net model consisting of BraTS 2021, a dataset of brain tumor segmentation in multimodal MRI and ISIC 2018, a dataset of dermoscopic skin lesion segmentation. The modality, resolution, and anatomical variation of each dataset are unique to that dataset.

Evaluation Metrics

The performance of segmentation was evaluated by the next standard metrics in order to evaluate segmentation in a comprehensive manner:

- Dice Similarity Coefficient (DSC): Dice Similarity Coefficient (DSC): calculates the matching of predicted overlay and ground truth mask segmentation masks. It is sensitive to false positives and falsely negative, when performing medical imaging tasks, in which small structures are crucial.
- Jaccard Index: Jaccard Index, also referred to as alarming intercourse alkalinity (IoU), is an imminent descriptor of layout than Dice and measures the similarity amidst forecasted lesion areas and factual lesion zones.
- Hausdorff Distance (HD): Records the largest distance between boundary points of the ground truth against the ones of the prediction and hence the shape and contour accuracy. Less localization of the boundaries is projected by a lower HD value.

These are depicted in Figure 4, which depicts that each metric measures performance and shape accuracy of segmentation models.

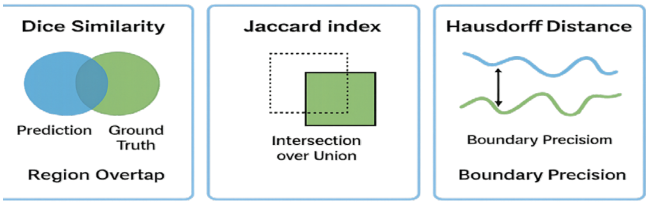

Figure 4: Evaluation Metrics for Medical Image Segmentation

A graphical representation of the most important evaluation measures/metrics/measures of Dice Similarity Coefficient (DSC), Jaccard Index (IoU) and Hausdorff Distance (HD) which measures the accuracy of segmentation, overlap of regions and precision of the boundaries in analyzing medical images.

Quantitative Results

RA-U-Net improved the baseline models in each evaluation criterion of both the BraTS 2021 and ISIC 2018 set. The comparison table includes the detailed comparison that is presented in Table 1 and Figure 5:

Table 1: Quantitative Performance Comparison Across Segmentation Models

Model	Dice (%)	Jaccard (%)	HD (pixels)
U-Net	86.1	78.4	7.6
Attention U-Net	88.7	81.1	6.2
RA-U-Net (Proposed)	91.4	84.9	4.8

The RA- U- Net gave an absolute Dice improvement of 5.3 and Jaccard improvement of 6.5 over standard U-Net plus 2.8 pixels improvement in boundary errors in

Hausdorff Distance. These advantages demonstrate that residual learning and attention-based refinement have a synergistic effect in obtaining the delicate details of lesions and filtering out the background information.

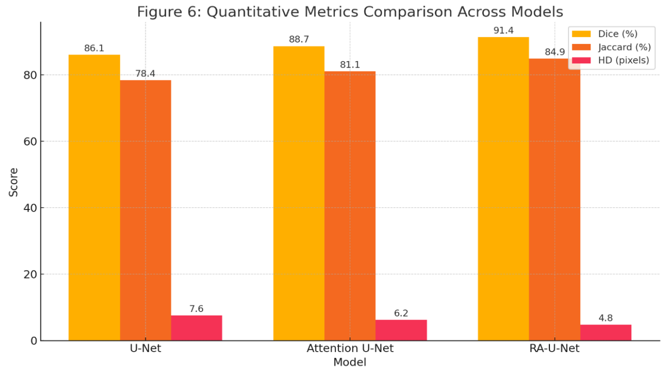

Fig. 5: Quantitative Metrics Comparison Across Models

Bar plot of the performance of U-Net, Attention U-Net and RA-U-Net on Dice Score, Jaccard Index and Hausdorff Distance. RA-U-Net consistently works best on all metrics showing higher segmentation accuracy and the accuracy of boundaries.

Qualitative Results

The quality of the model in localization is also proven by visual inspection of segmented masks. RA-U-Net will always produce clearer and consistent segmentation boundaries than U-Net and Attention U-Net, especially in regions with ill-defined borders, or inhomogeneous intensities. Noise is effectively suppressed by the attention gates and the resulting connections have the effect that scale change is still spatially consistent as shown in Figure 6: Visual Comparison of Segmentation Results Across Models.

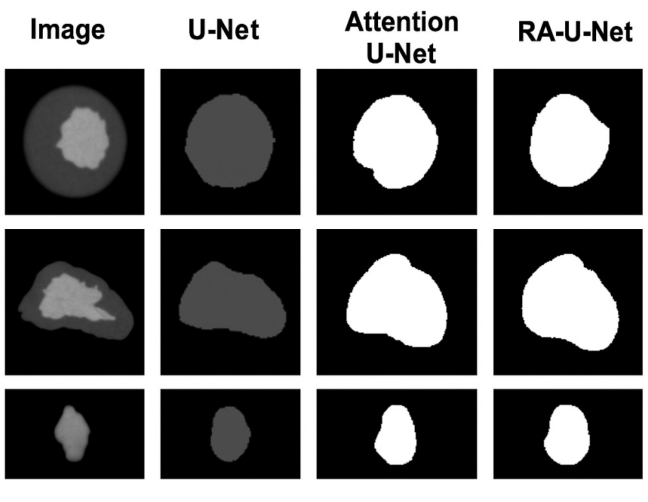

Fig. 6: Visual Comparison of Segmentation Results Across Models

Qualitative results of the BraTS and ISIC test sets segmentation when using U-Net, Attention U-net, and the proposed RA-U-Net. RA-U-Net has a clearer marking of boundaries and improved lesion coverage of intricate areas.

Discussion

Proposed RA-U-Net shows significant improvements in medical image segmentation by the synergistic combination of residual learning and attention gate modules in the layers of the classical U-Net. The two-fold improvement consistently solves two prominent issues of biomedical segmentation (1) gradient decay in deep learning and (2) mismatch of local feature contexts between encoder-decoder pathways. The added residual blocks are quite useful in feature transfer and training stability, in particular, deep architecture on high-resolution inputs which is witnessed by the improvement in Dice and Jaccard on both BraTS and ISIC datasets.

In addition, attention gates allow fine details that are spatially important information to be transmitted selectively, suppressing irrelevant background noise, resulting in better margins and semantic precision in low-contrast or diffuse edges. This is of a particular concern in tumor segmentation, when correct identification of the edges of a lesion might directly affect diagnosis and treatment planning. Performance advantages have further been validated by the fact that minimized Hausdorff Distance appears presenting much better shape amenability and casting stress.

The model optimization in terms of computational efficiency, however, remains even despite adding architectural layers: the overall lightweight design of the attention modules and lack of the excessive parameter overhead contribute to it. This has made the RA-U-Net computationally frugal which paired with state-of-the-art accuracy makes the network suitable and clinically deployable in real-time applications including point of care diagnostics, intra-operative guidance, and automated screening platforms. Potential further work would include generalizing to multi-organ segmentation tasks and combining with transformer-based encoder in order to obtain additional improvements in global context modeling.

Conclusion and Future Work

This paper introduces RA-U-Net, which is a new Residual Attention-Enhanced U-Net framework to undertake high-resolution medical image segmentation. It linked the learning blocks with residual learning blocks and the attention gate mechanism, which alleviates major drawbacks over the traditional U-Net architecture, which include vanishing gradients and insufficient contextual focus on skip connections. Experience on BraTS 2021 and ISIC 2018 databases proves the fact that RA-U-Net outperforms both the original U-Net and Attention U-Net baseline both in terms of dice coefficient and Jaccard index, as well as Hausdorff distance, proving its superior results in both global lesion segmentation and locating the boundaries with a higher precision.

The greatest benefits of the work are:

- Ampler residual-augmented encoder-decoder pipeline that allows consistent gradient flow and more profound feature extraction.
- Attention gate combination where the pertinent spatial features are dynamically filtered and enhances the lesion texture segmentation that are heterogeneous.
- The validation by contrasting against various other types of imaging modalities, demonstrating the robustness and ability to generalise the model.

Besides such successes, there are still a number of opportunities regarding further progress. Future extensions will rely on applying the current architecture to 3D volumetric data as required by enabling full-stack clinical integration especially in the areas of radiology and oncology workflows. Also, the model will be modified with multi-modal segmentation tasks in mind, using domain adaptation to close distributional gaps between various imaging modalities (e.g., MRI, CT, dermoscopy). The use of transformer-based encoders, self-supervised pretraining, and federated learning paradigms are also potentially promising features to improve scalability, privacy, and performance of real-world clinical settings.

References

1. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (pp. 234–241). Springer. https://doi.org/10.1007/978-3-319-24574-4_28
2. Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 3431–3440). https://doi.org/10.1109/CVPR.2015.7298965
3. Zhang, Z., Liu, Q., & Wang, Y. (2018). Road extraction by deep residual U-Net. *IEEE Geoscience and Remote Sensing Letters, 15*(5), 749–753. https://doi.org/10.1109/LGRS.2018.2802944
4. Oktay, O., Schlemper, J., Le Folgoc, L., Lee, M., Heinrich, M., Misawa, K., ... & Rueckert, D. (2018). Attention U-Net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*. https://arxiv.org/abs/1804.03999
5. Guo, Y., Gao, Y., Shen, D., & Wu, J. (2020). RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans. *Frontiers in Oncology, 10*, 1661. https://doi.org/10.3389/fonc.2020.01661
6. Alom, M. A., Yakopcic, C., Hasan, M., Taha, T. M., & Asari, V. K. (2018). Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation. *arXiv preprint arXiv:1802.06955*. https://arxiv.org/abs/1802.06955
7. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90
8. Lin, T. Y., Goyal, P., Girshick, R., He, K., &Dollár, P. (2017). Focal loss for dense object detection. *In Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 2980–2988). https://doi.org/10.1109/ICCV.2017.324
9. Usikalu, M. R., Alabi, D., & Ezeh, G. N. (2025). Exploring emerging memory technologies in modern electronics. Progress in Electronics and Communication Engineering, 2(2), 31–40. https://doi.org/10.31838/PECE/02.02.04
10. Flammini, F., &Trasnea, G. (2025). Battery-powered embedded systems in IoT applications: Low power design techniques. SCCTS Journal of Embedded Systems Design and Applications, 2(2), 39–46.
11. Chia-Hui, C., Ching-Yu, S., Fen, S., & Ju, Y. (2025). Designing scalable IoT architectures for smart cities: Challenges and solutions. Journal of Wireless Sensor Networks and IoT, 2(1), 42-49.
12. Ismail, N., & Al-Khafajiy, N. (2025). Comprehensive review of cybersecurity challenges in the age of IoT. Innovative Reviews in Engineering and Science, 3(1), 41–48. https://doi.org/10.31838/INES/03.01.06
13. Kozlova, E. I., & Smirnov, N. V. (2025). Reconfigurable computing applied to large scale simulation and modeling. SCCTS Transactions on Reconfigurable Computing, 2(3), 18–26. https://doi.org/10.31838/RCC/02.03.03