2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Dual Encoder Attention U-net for nuclei segmentation

Abhishek Vahadane, Atheeth B, Shantanu Majumdar
Rakuten Institute of Technology India, Rakuten, Inc.

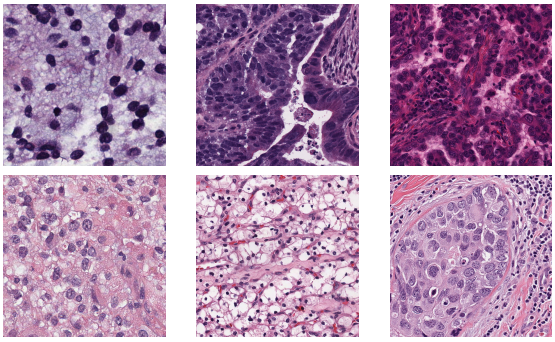
***Abstract*— Nuclei segmentation in whole slide images (WSIs) stained with Hematoxylin and Eosin (H&E) dye, is a key step in computational pathology which aims to automate the laborious process of manual counting and segmentation. Nuclei segmentation is a challenging problem that involves challenges such as touching nuclei resolution, small-sized nuclei, size, and shape variations. With the advent of deep learning, convolution neural networks (CNNs) have shown a powerful ability to extract effective representations from microscopic H&E images. We propose a novel dual encoder Attention U-net (DEAU) deep learning architecture and pseudo hard attention gating mechanism, to enhance the attention to target instances. We added a new secondary encoder to the attention U-net to capture the best attention for a given input. Since H captures nuclei information, we propose a stain-separated H channel as input to the secondary encoder. The role of the secondary encoder is to transform attention prior to different spatial resolutions while learning significant attention information. The proposed DEAU performance was evaluated on three publicly available H&E data sets for nuclei segmentation from different research groups. Experimental results show that our approach outperforms other attention-based approaches for nuclei segmentation.**

Keywords– Dual Encoder, Nuclei segmentation, Attention.

I. Introduction

Automated detection of cell nuclei in H&E [1] stained whole slide images (WSIs) is a necessary step in digital pathology since manual examination of WSIs is tedious and prone to inter and intra-observer variations [2]. Fig. 1 shows the inherent variations in size, opacity, and color of nuclei which pose challenges in nuclei segmentation. Several different techniques were proposed. It includes techniques such as watershed segmentation, active contour, level-sets, snake energy optimization, morphological processing and their variants [3], [4], [5], [6], [7], [8], [9], [10]. However, these techniques cannot generalize on the challenges discussed in Fig. 1.

In recent times, deep learning techniques have been shown to outperform energy-based models for the problem of nuclei segmentation. CNN3 [11] proposed a three-class pixel classification using a convolutional neural network (CNN) where each pixel in the input image has a predicted probability score for three classes: nucleus inside, outside, and boundary. The nucleus inside and boundary probability maps were further used simultaneously to obtain robust nuclei segmentation. Inspired from fully convolutional Networks, FCN8 [12] and U-Net [13] became popular for many medical image segmentation problems. U-net consists of an encoder-decoder setup with skip connections from the encoder to the decoder. The skip connections enable access to low-high level features from the encoder to the decoder. However, it remains difficult to prune the skip connections to capture features related to small objects like nuclei. To tackle this problem with small objects, attention gates(AGs) were integrated with standard CNN models[14].



(a)Opacity (b)Size (c) Color

Fig. 1. The inherent variations exhibited by nuclei in H&E images.

Attention gates were initially used in sequence modeling in tasks such as Natural Language Processing and machine translation. In machine translation [15], attention is used to weigh each of the input sequence's hidden states and provide it as an additional input to each output sequence's hidden state. Grid-based attention maps have gained popularity in computer vision to improve model performance and are categorized as either hard and soft. Hard attention [16] is non-differentiable and relies on highly intractable techniques for parameter update while soft attention like the one used in [14] has no explicit way to control the regions to which it pays attention. When dealing with small objects like nuclei, such unconstrained attention maps can result in a large number of false negatives [17]. Moreover, in backpropagation, the gradients with respect to the output of attention gates are again multiplied by the attention coefficients. In theory, it is supposed to allow shallow layers to receive parameter updates from only relevant features, but poor attention maps can harm the overall performance of the model. We propose to improve the attention capturing capacity of FCNs to boost nuclei segmentation performance.

Our goal is to tackle the challenges depicted in Fig.1 using a deep learning-based approach. The core idea of grid-based attention is to learn attention coefficients that identify salient image regions, objects, and spatial arrangements of different objects which reduce the total network loss. Attention mechanism like AGs prunes the feature responses so that relevant activation to the specific task is preserved. We propose a novel network with AGs to improvise the

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

attention mechanism of the entire network. As attention U-net was demonstrated to improve performance through AGs, we demonstrate the capability of proposed work on attention U-net. In this paper, our contributions are as follows:

1) Proposed a dual encoder architecture that encodes attention prior information. We also suggest a way to generate attention prior to input H&E images.
2) A novel attention skip module (ASM) that utilizes both attention prior and input feature maps to enhance segmentation performance.

Extensive experiments show that the proposed architecture improves the attention-catching capability and performance of the network.

II. Methodology

In this section, we present an end-to-end deep learning framework to segment nuclei instances accurately, given H&E histology images. The block diagram of the framework is shown in Fig 2. It consists of pre-processing step to generate the attention prior from H&E histology images, a deep learning architecture that incorporates a novel attention mechanism, and a post-processing set-up to produce robust nuclei segmentation. In order to separate touching and overlapping nuclei, we employ a contour based approach [11], [18] to predict the nuclei and its corresponding boundary probability maps. In the post-processing set-up, the predicted nuclei and boundary probability maps are used to further refine the nuclei segmentation.

Attention Prior: We proposed to use the Hematoxylin (H) channel optical density in H&E image as attention prior as it stains the nuclei. Sparse non-negative matrix factorization (SNMF) [19] method was used to deconvolve and normalize the H&E stained WSIs into H and E optical density maps, of which H map can be used for generating attention prior.

Dual encoder Attention U-net (DEAU): As shown in Fig. 3, we incorporated a novel Attention Encoding Path (AEP) to a conventional U-Net architecture, which takes the attention prior as input and learns meaningful features for segmentation. The dimensions of the new encoding path is same as the other trivial encoding path. The trivial encoding path is given an H&E input image. The AEP has attention prior as the input. The feature maps obtained from the trivial encoding path and AEP are fed to the attention skip module (ASM) at different depth of the network. The proposed Attention Skip Module (ASM) is shown in the Fig. 4. It is a modification of the gating mechanism proposed in [14]. At each spatial resolution of the skip connection, the module accepts the processed attention feature map $g^l \in R^{F_H}$ and H&E feature map $x^l \in R^{F_{HE}}$ feature map as its two inputs. These vectors are translated to an intermediate dimension F_{int} by 1×1 convolution with kernels W_x and W_g. This is followed by element wise addition, non linear transformation through ReLU function σ_1 and another 1×1 convolution operation with kernel (W_{int}), before passing the output to the sigmoid function σ_2. This generates attention coefficients $\alpha^l \in [0, 1]$. The final output of the module $\hat{x}^l$ is obtained by element-wise multiplication of α^l with x^l, followed by a final 1×1 convolution. As each of the convolution operations have associated parameters that are updated during back propagation by the gradients, the attention mechanism can be called as pseudo hard gating. At a skip connection l, the attention coefficient is given by Equation 1 where b_1 and b_2 are bias terms.

$$\alpha^l = \sigma_2 \left(W_{int}^T \left(\sigma_1 \left(W_x^T \, x^l + W_g^T \, g^l + b_1 \right) \right) + b_2 \right) \quad (1)$$

Post-Processing: The raw output of the DEAU consists of nuclei prediction $I_n \in R^{H \times W}$, and boundary prediction $I_b \in R^{H \times W}$. We use the post-processing logic of [11] to semantically segment the nuclei as well as separate touching nuclei. In the first step, I_n and I_b are thresholded using empirically determined thresholds 0.45 and 0.3 respectively. The binary boundary map $\hat{I}_b$, is subtracted from binary nuclei map $\hat{I}_n$, resulting in segregated nuclei instance map z_i.Then, we generate an energy landscape in the form of a distance map d for each individual connected component instance. where the distance of each pixel to background is calculated. To generate markers I_m of nuclei, we further erode z_i. Utilizing the distance map and the markers of segregated nuclei, we employ marker-controlled watershed [19] and get an N-array mask of nuclei instances.

III. Experiments and Results

A. Datasets

Few research groups have already open-sourced nuclei segmentation datasets such as Kumar et.al [11], CoNSep [20] and CPM-17 [21]. Kumar et.al [11] dataset contains 21,000 annotated nuclei from different organ tissue H&E images such as prostate, colon, breast, kidney, prostate, liver, bladder, and stomach. CoNSep [20] is focused on colon tissue H&E stained images, however, images were selected owing to variability and diversity of tissue structure. The CPM-17 [21] was made available as a part of the MICCAI2017 challenge and has tissue images of patients with Non-Small Cell Lung Cancer (NSCLC), Head and Neck Squamous Cell Carcinoma (HNSCC), Glioblastoma Multiforme (GBM), and Lower Grade Glioma (LGG) tumors. We validate the proposed DEAU on these three datasets and also compare DEAU with U-net based approaches.

B. Implementation Details

Fixed-size patches of spatial dimension 256×256 were randomly extracted from each data-set after performing relevant zero paddings. To increase the data samples and variability, we dynamically augment the patches. Augmentation includes random horizontal flip, Gaussian blur with unit sigma, median blur with a kernel size of 3×3 and random rotation between $60°$ to $120°$. Proposed DEAU framework was implemented in the open-source deep learning library PyTorch 1.3.1. The parameters of the model were initialized as sampled values drawn from a normal distribution with zero mean and unit standard deviation. The number of training epochs for each experiment was fixed to 80 and we chose Adam optimizer. The initial learning rate was 10^{-04}, which

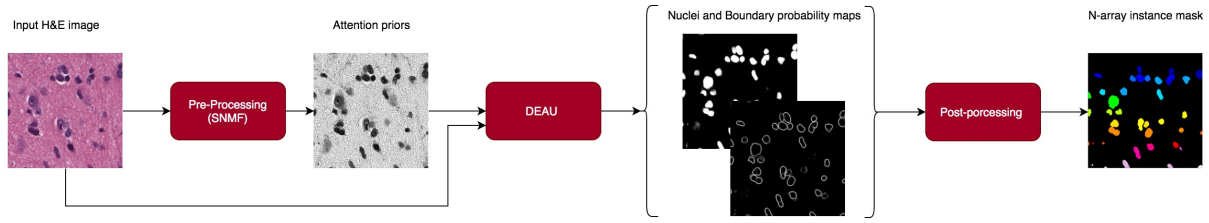


Fig. 2. The complete framework of the proposed nuclei segmentation approach.

TABLE I

RESULTS OF COMPARATIVE EXPERIMENTS CARRIED OUT ON KUMAR, CONSEP AND CPM-17 DATA SETS.

	Kumar					CoNSep					CPM-17				
Methods	**Dice**	**AJI**	**DQ**	**SQ**	**PQ**	**Dice**	**AJI**	**DQ**	**SQ**	**PQ**	**Dice**	**AJI**	**DQ**	**SQ**	**PQ**
Cell Profiler [22]	0.623	0.366	0.423	0.704	0.300	0.434	0.202	0.249	0.705	0.179	0.570	0.338	0.368	0.702	0.261
QuPath [23]	0.698	0.432	0.522	0.679	0.351	0.588	0.249	0.216	0.641	0.151	0.693	0.398	0.320	0.717	0.230
FCN8 [12]	0.797	0.281	0.434	0.714	0.312	0.756	0.123	0.239	0.682	0.163	0.840	0.397	0.575	0.750	0.435
U-Net [13]	0.758	0.556	0.691	0.690	0.478	0.724	0.482	0.488	0.671	0.328	0.813	0.643	0.778	0.734	0.578
DIST [18]	0.789	0.559	0.601	0.732	0.443	0.804	0.502	0.544	0.728	0.398	0.826	0.616	0.663	0.754	0.504
CNN3 [11]	0.762	0.508	-	-	-	-	-	-	-	-	-	-	-	-	-
Attention U-Net [14]	0.786	0.416	0.459	0.690	0.319	**0.831**	0.494	0.538	0.743	0.402	0.846	0.592	0.746	0.781	0.587
Micro-Net [24]	0.797	0.560	0.692	0.747	0.519	0.794	0.527	0.600	0.745	0.449	**0.857**	0.668	0.836	0.788	0.661
Hover-net [20]	**0.826**	0.618	0.770	0.773	0.597	**0.853**	0.571	0.702	0.778	0.547	**0.869**	0.705	0.854	0.814	0.697
Dual Encoder	**0.810**	0.516	0.594	0.737	0.440	**0.822**	0.485	0.492	0.744	0.368	**0.857**	0.627	0.773	0.787	0.609

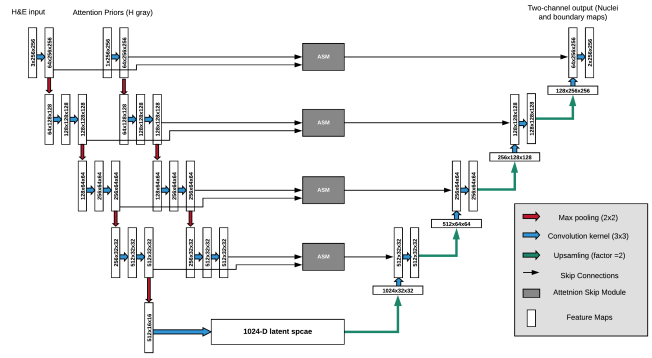


Fig. 3. Schematic of the proposed DEAU with the Attention Encoding Path (AEP) and the novel attention skip module.

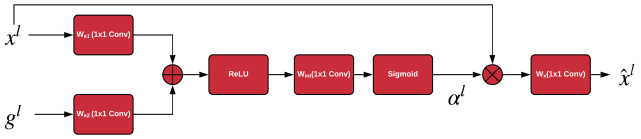


Fig. 4. Schematic of the proposed Attention Skip module (ASM). Input features x^l are scaled with attention coefficients α^l computed in ASM.

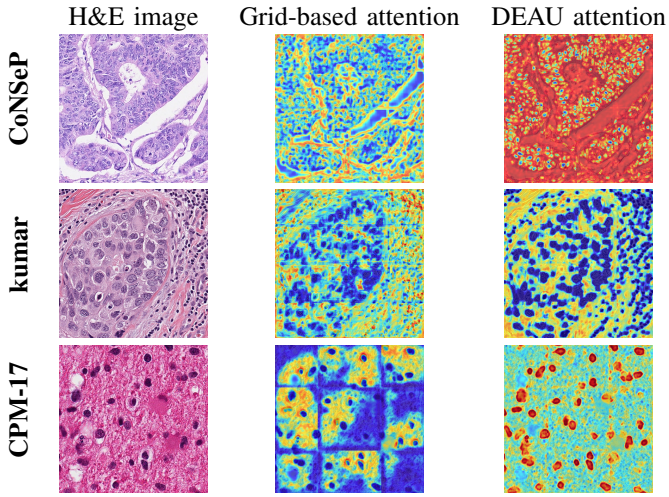


Fig. 5. The attention maps[2] at spatial resolutions 32×32 and 64×64, scaled up for visualization. While the actual weight-age given to different regions in the slide images are irrelevant, the uniformity with which they are assigned is important. It can be seen that in the case of pseudo-hard attention maps the weight-age assigned to nuclear regions are near homogeneous, as compared to their counterparts in grid-based approach.

reduces by a factor of 10^{-1} when the validation dice score does not change across four epochs.

C. Evaluation metrics

We adopt the typical metrics uses to measure segmentation performance. Dice coefficient (DICE) between two sets X and Y is defined as $\frac{2\times(|X\cap Y|)}{(|X|+|Y|)}$ Aggregated Jacquard Index (AJI) [11] on the other hand computes the ratio of an aggregated intersection cardinality and an aggregated union cardinality between X and Y. Detection Quality (DQ)is the object level F-1 score given by, $\frac{2\times(|TP|)}{(2|TP|+|FP|+|FN|)}$. For every (x, y) pair in (ground truth, prediction), Segmentation Quality (SQ) is given by $\frac{\sum_{(x,y)\in TP} IoU(x,y)}{|TP|}$. Panoptic quality (PQ) [25] is the final metric used, and it is the product of DQ and SQ.

D. Baseline Methods

FCN8: We use the popular FCN8 [12] as a baseline to predict nuclei and boundary probability maps. The predictions are fed to the post-processing set-up of [20] to generate N-array masks. **U-Net (2-class)**: Like the previ-

ous approach, U-Net architecture [13] predicts 2 output classes (nuclei,boundary). To generate instance maps from raw prediction we follow the same post-processing procedure as above. **CNN3**: The approach in [11] carries out pixel-level classification of input H&E image. Among the deep learning approaches compared, this is the only non fully convolutional network. **Attn U-Net**: An extension to the U-Net was proposed in [14]. The changes include a modification to the skip connections, to incorporate a grid-based attention mechanism. We also compare with other methods and popular image analysis software such as Cell Profiler [22] and QuPath [23].

E. Experimental Results

The quantitative comparison of DEAU with contemporary architecture is shown in Table I. We trained Attention U-net and DEAU on three data-sets discussed previously. The results of U-net, FCN8, DIST, Micro-Net, and Hover-net were taken from [20]. As shown in Table I, proposed DEAU outperforms attention U-net in almost all the metrics which proves the effectiveness in attention mechanism brought by DEAU. DEAU performs equally with other top performers such as Micro-Net and DIST when we compare the Dice. We need to improve the post-processing of our approach to further refine the other instance segmentation metrics.

We showed the qualitative results in Fig. 5 (columns 2 and 3). Since DEAU and attention U-net generate attention maps as metadata, we compare attention mechanisms of both methods by visualizing the attention maps generated by each of them. As shown in Fig. 5, DEAU based attention mechanism shows accurate localization of nuclei. However, attention U-net shows poor attention and sometimes confusing. As an example, in the sample from Kumar (2nd row), on the right top region, grid-based attention shows high attention to nuclei, however, in the center, the attention was poor on nuclei. DEAU was consistent where it has a high attention to nuclei in the entire image.

IV. Conclusion

In this paper, we presented a novel attention-based deep learning approach to segment nuclei. The method was build upon a novel secondary encoder to constrain the DEAU to focus on only relevant regions of H&E image such as nuclei. The input to the secondary encoder, stain separated H channel, help DEAU not to look for all over the image. We conclude that the addition of a secondary encoder and logical attention prior significantly helped the network to learn a better representation in the latent space. This also enabled the decoder to produce improved probability maps to segment nuclei. Quantitative comparison with other contemporary approaches shows that the method achieves a higher dice score. The other metrics such as AJI, DQ, SQ, and PQ depend upon the accuracy of post-processing techniques. In the future, we want to tune our post-processing further to improve the performance. We analyzed through visual comparison, that the attention maps generated by our method are finer and more consistent than grid-based attention approach.

References

[1] Humayun Irshad et.al. Methods for nuclei detection, segmentation, and classification in digital histopathology: a review—current status and future potential.

[2] Henry K Su et.al. Inter-observer variation in the pathologic identification of minimal extrathyroidal extension in papillary thyroid carcinoma. *Thyroid*, 26(4):512–517, 2016.

[3] Mitko Veta et.al. Automatic nuclei segmentation in h&e stained breast cancer histopathology images. *PloS one*, 8(7), 2013.

[4] Jierong Cheng et.al. Segmentation of clustered nuclei with shape markers and marking function. *IEEE Transactions on Biomedical Engineering*, 56(3):741–748, 2008.

[5] Hui Kong et.al. Partitioning histopathological images: an integrated framework for supervised color-texture segmentation and cell splitting. *IEEE transactions on medical imaging*, 30(9):1661–1677, 2011.

[6] Xiaodong Yang et.al. Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and kalman filter in time-lapse microscopy. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 53(11):2405–2414, 2006.

[7] Sahirzeeshan et.al. An integrated region, boundary, shape-based active contour for multiple object overlap resolution in histological imagery. *IEEE transactions on medical imaging*, 31(7):1448, 2012.

[8] Stephan Wienert et.al. Detection and segmentation of cell nuclei in virtual microscopy images: a minimum-model approach. *Scientific reports*, 2:503, 2012.

[9] Antonio LaTorre et.al. Segmentation of neuronal nuclei based on clump splitting and a two-step binarization of images. *Expert Systems with Applications*, 40(16):6521–6530, 2013.

[10] Miao Liao et.al. Automatic segmentation for cell images based on bottleneck detection and ellipse fitting. *Neurocomputing*, 173:615–622, 2016.

[11] Neeraj Kumar et.al. A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE transactions on medical imaging*, 36(7):1550–1560, 2017.

[12] Jonathan Long et.al. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[13] Olaf Ronneberger et.al. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[14] Ozan Oktay et.al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.

[15] Ashish Vaswani et.al. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.

[16] Volodymyr Mnih et.al. Recurrent models of visual attention. In *Advances in neural information processing systems*, page 2204, 2014.

[17] Inwan Yoo et.al. Pseudoedgenet: Nuclei segmentation only with point annotations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 731–739. Springer, 2019.

[18] Peter Naylor et.al. Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE transactions on medical imaging*, 38(2):448–459, 2018.

[19] Abhishek Vahadane et.al. Towards generalized nuclear segmentation in histological images. In *IEEE International Conference on BioInformatics and BioEngineering*, pages 1–4. IEEE, 2013.

[20] Simon Graham et.al. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Analysis*, 58:101563, 2019.

[21] Quoc Dang Vu et.al. Methods for segmentation and classification of digital microscopy tissue images. *Frontiers in bioengineering and biotechnology*, 7, 2019.

[22] Anne Carpenter et.al. Cellprofiler: image analysis software for identifying and quantifying cell phenotypes. *Genome biology*, 7(10), 2006.

[23] Peter Bankhead et.al. Qupath: Open source software for digital pathology image analysis. *Scientific reports*, 7(1):1–7, 2017.

[24] Shan E Ahmed Raza et.al. Micro-net: A unified model for segmentation of various objects in microscopy images.

[25] Alexander Kirillov et.al. Panoptic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9404–9413, 2019.