

Efficient Multi-Organ Multi-Center Cell Nuclei Segmentation Method Based on Deep Learnable Aggregation Network



Loay Hassan¹, Adel Saleh², Mohamed Abdel-Nasser^{1,3*}, Osama A. Omer¹, Domenec Puig³

¹Electrical Engineering Department, Aswan University, Aswan 81528, Egypt

²Gaist Solutions Ltd, Skipton BD23 2TZ, UK

³Department of Computer Engineering and Mathematics, University Rovira i Virgili, Tarragona 43007, Spain

Corresponding Author Email: egnaser@gmail.com

<https://doi.org/10.18280/ts.380312>

ABSTRACT

Received: 18 March 2021

Accepted: 5 June 2021

Keywords:

computer-aided diagnosis, deep learning, digital pathology, nuclei segmentation, whole slide imaging

Automated cell nuclei delineation in whole-slide imaging (WSI) is a fundamental step for many tasks like cancer cell recognition, cancer grading, and cancer subtype classification. Although numerous computational methods have been proposed for segmenting nuclei in WSI images based on image processing and deep learning, existing approaches face major challenges such as color variation due to the use of different stains, the various structures of cell nuclei, and the overlapping and clumped cell nuclei. To circumvent these challenges in this article, we propose an efficient and accurate cell nuclei segmentation method based on deep learning, in which a set of accurate individual cell nuclei segmentation models are developed to predict rough segmentation masks, and then a learnable aggregation network (LANet) is used to predict the final nuclei masks. Besides, we develop cell nuclei segmentation software (with a graphical user interface—GUI) that includes the proposed method and other deep-learning-based cell nuclei segmentation methods. A challenging WSI dataset collected from different centers and organs is used to demonstrate the efficiency of our method. The experimental results reveal that our method obtains a competitive performance compared to the existing approaches in terms of the aggregated Jaccard index (AJI=89.25%) and F1-score (F1=73.02%). The developed nuclei segmentation software can be downloaded from <https://github.com/loaysh2010/Cell-Nuclei-Segmentation-GUI-Application>.

1. INTRODUCTION

Digital pathology along with digital image analysis tools are gaining strength in clinical use day-to-day, thanks to the significant improvements in equipment, computational technology, and storage. In short, whole-slide imaging (WSI) stands for the process of digitalizing the glass histology slides and capturing high-resolution images using scanning devices. WSI technology has allowed pathologists to perform robust analysis, processing, and management of thousands of tissue biopsies taken from cancer patients [1]. Considering the large size of the WSI images and the rich information they contain, pathologists do spend a long time analyzing such images manually. Over the last years, the employment of computerized approaches is rapidly evolving with different potential digital pathology applications, such as cell nuclei segmentation, cell classification, counting cancer cells, and cancer prognosis.

Cell nuclei segmentation is a key process in the field of WSI image analysis. In particular, accurate and robust nuclei segmentation tools are required to automatically extract and interpret sub-cellular morphologic and shape information in WSI images. Given the segmented cell nuclei, different features and descriptors like the number and shape of cell nuclei are used for determining cancer types, cancer grading, and cancer prognosis [2]. In this context, the term ‘cell segmentation’ stands for classifying each pixel in the WSI

image as cell nuclei and non-nuclei pixel. Hence, each nucleus can be extracted from the image and made available for further analysis. The automated segmentation of cell nuclei is still a challenging process because each medical center uses different stains (i.e., color variation) to produce the WSI images, as well as the presence of adjacent and overlapping cells in WSI images.

In the last decade, deep learning technologies have obtained high performance in different image segmentation and classification tasks [3-5], thanks to the robust feature representation of convolutional neural networks (CNNs). In the field of medical and biological-image analysis, deep learning has achieved successful results with challenging and complex segmentation tasks [6-8]. For instance, mask-RCNN was used by Arai and Kapoor [9] to segment nuclei in a fully automatic fashion. Mahmood et al. [10] used an unpaired GAN architecture to synthesize WSI images with perfect nuclei delineation. Then, they used the synthesized WSI images, real WSI images, spectral normalization, and gradient penalty to train a conditional generative adversarial network (cGAN) for cell nuclei segmentation. However, some factors like the clumped cell nuclei and overlapping cell nuclei as well as the color variation in WSI images (various staining routines) and the ambiguous boundary between different cell nuclei limit the performance of the existing deep learning-based approaches.

To circumvent these challenges, we propose an accurate and efficient deep learning-based method for segmenting cell

nuclei in WSI images acquired from multiple organ and multiple centers. Specifically, we use deep CNN architectures to develop accurate individual cell nuclei segmentation models (ISs). The input WSI images are fed into the trained ISs to predict rough cell nuclei masks, and then a deep learnable aggregation network called LANet is used to aggregate the predictions of ISs guided by the input WSI images.

The key contributions of this article are highlighted below:

- We propose an accurate and efficient nuclei segmentation method, in which a deep aggregation network (LANet) is constructed to fuse the predictions of accurate individual cell nuclei segmentation models to obtain precise delineation.
- We present a detailed analysis for the proposed cell nuclei segmentation method and comparisons with the existing cell nuclei segmentation methods (FCN8s [11], UNet [8], UNet++ [12], SegNet [13], RIC-UNet [14], DIST [15] and cGANs [10]) and existing cell nuclei segmentation software (ImageJ [16]) using a challenging multiple organ and multiple center dataset.
- We develop a new nuclei segmentation software (with a graphical user interface—GUI) that includes the proposed method as well as deep-learning-based nuclei segmentation methods. The software can be downloaded from the following link <https://github.com/loaysh2010/Cell-Nuclei-Segmentation-GUI-Application>.

This paper includes 5 sections. Section 2 presents and discusses the state-of-art methods and highlights their limitations. The proposed cell nuclei segmentation method and the implementation steps are presented in Section 3. The experimental results comparisons with the existing cell nuclei segmentation methods and discussions are provided in Section 4. The conclusions and the future work are given in Section 5.

2. RELATED WORK

In the last years, deep learning technologies have made breakthroughs in the analysis of digital images. One of the key advantages of deep CNNs is their ability to generate an accurate representation of the images. Unlike hand-crafted methods, deep CNNs can be trained on an end-to-end fashion to learn descriptive representations from the input images. Such merits have increased the interest of researchers and developers investigating the applicability of DCNNs-based methods with the problem of cell segmentation. Several forms of the neural network have been derived from CNNs for cell segmentation.

Zhou et al. [12] proposed the UNet++ model, a deep learning-based model for semantic (instance) segmentation, in which they combined UNets of varying depths into one architecture. The decoders of the UNets are densely connected via a skip connection mechanism. Besides, they used a deep supervision technique to prune the UNet++ model. This pruning technique could accelerate inference time of UNet++ model without degrading its performance. Zhou et al. [12] demonstrated that UNet++ outperformed many medical image segmentation models like UNet.

Pan et al. [17] introduced a deep learning-based cell nuclei segmentation called AS-UNet. It should be noted that the main difference between AS-UNet and UNet is that AS-UNet contains atrous depth-wise separable convolution blocks. The key components of AS-UNet are (1) an encoder module to

extract high-level semantic information from the input images, 2) a decoder module to restore the spatial information of the input images and produce the segmentation masks, and 3) an atrous convolution module (cascaded and parallel atrous convolutions). In AS-UNet, the atrous convolution module is added between the encoding and decoding module. As claimed by Pan et al. [17], the use of the atrous convolution can improve the receptive fields of the segmentation model without increasing the computational complexity. AS-UNet achieved good performance on two cell nuclei segmentation datasets (MoNuSeg [18] and BNS [19]). It obtained an F1-score of 87.35% with MoNuSeg and 86.97% with BNS.

Zeng et al. [14] proposed the RIC-UNet model for nuclei segmentation. They integrated the residual learning, channel attention, and multiscale approaches with the UNet architecture to segment nuclei precisely. They integrated residual connections [5] in the down-sampling path of the basic architecture of UNet. The use of residual connections helps better represent image features with computational efficiency of the inception module [20] while incorporating multiscale features with different kernel sizes. Besides, they integrated a channel attention mechanism [21] in the up-sampling path of the segmentation model to manipulate the heterogeneity of cell nuclei appearances in the WSI images. Zeng et al. also compared their method with well-known image analysis software, namely cell profile [22] and Fiji (a package based on ImageJ) [16]. They obtained an F1-score of 82.78% and AJI score of 56.35% using MoNuSeg [18] dataset.

Graham et al. [23] presented a deep learning-based method called Hover-Net for segmenting and classifying nuclear instances in WSI images simultaneously. It should be noted that Hover-Net is designed to predict horizontal and vertical distances of nuclear pixels from their centers of mass. Hover-Net also contains an up-sampling branch to recognize the nuclear class of each segmented instance. They demonstrated the efficacy of Hover-Net with six multiple tissue histopathological image datasets and presented a detailed comparative study. They obtained an AJI score of 61.80% with the MoNuSeg [18] dataset. Qu et al. [24] used partial points annotation in WSI images to construct a weakly supervised deep learning-based cell nuclei segmentation model. Specifically, the Voronoi diagram and the k-means clustering algorithm are used to process the input WSI images and the shape prior of cell nuclei to produce coarse labels. Afterwards, the coarse labels are used to train a deep learning-based nuclei segmentation model. The dense conditional random field is utilized in the loss function of the segmentation model. Qu et al. demonstrated that their method achieved segmentation similar to the ones of fully supervised nuclei segmentation approaches with few annotated data.

To circumvent the problem of touching and overlapping cell nuclei, Naylor et al. [15] solved the segmentation process as a regression task by predicting the distance map of cell nuclei. Naylor et al. claimed that close or overlapping cell nuclei are delineated as one object, which yields poor segmentation results. Naylor et al. demonstrated that their approach can segment touching or overlapping nuclei and it outperformed other CNN pixel-based classification approaches that do not take the topological relationship among cells center pixels and those in their neighborhoods into account.

Furthermore, Naylor et al. [19] proposed a fully automated cell nuclei segmentation method based on three deep learning-based segmentation models, namely fully convolutional network (FCN), PangNet, and DeconvNet. Naylor et al. fused

the segmentation masks of the three cell nuclei segmentation models to obtain accurate segmentation results. However, this method often segments touching nuclei as one single object. They obtained an F1-score of 80.2% using their own published dataset (BNS). Graham and Rajpoot [25] presented a stain-aware and multi-scale deep CNN-based model called SAMS-Net. They used a weighted cost function sensitive to the intensity of the Haematoxylin (H) stain in the input WSI images, which makes SAMS-Net robust to Haematoxylin and eosin (H&E) stain variations. SAMS-Net incorporates information at different resolutions and employs skip connections to distinguish between separate nuclei.

Although the cell nuclei segmentation methods above mentioned have achieved a promising result, the clumped cell nuclei and overlapping cell nuclei as well as the color variation in WSI images (various H&E routines) limit the performance of the existing approaches. To circumvent these issues, we propose an efficient approach for segmenting nuclei in WSI images based on a set of accurate cell nuclei segmentation models and a deep CNN-based learnable aggregation network.

3. METHODOLOGY

Figure 1 presents an overview of the proposed cell nuclei segmentation method. As shown, the stains of all WSI images are normalized. The normalized WSI images are then fed into N accurate individual segmentation models (IS) to predict coarse nuclei segmentation masks (i.e., coarse labels). The coarse nuclei segmentation masks as well as the original RGB input image are concatenated and inputted into a deep learnable CNN-based aggregation network called LANet to produce the final nuclei segmentation mask. Below, we explain each step of the proposed method in details.

3.1 Stain normalization

Stain normalization is a fundamental step to segment nuclei in multi-center and multi-organ WSI images to reduce the color variation and obtain a better color consistency before feeding the images into the segmentation methods. In this study, we employ the stain normalization method [26], in which one image is selected as the target and all other images are converted to its color space. It should be noted that the inputs and outputs of the stain normalization stage are images, where all WSI images are transformed to the color space of a predefined target image. The stain normalization can be solved as an optimization problem as follows [26]:

$$\min_{W, H} \frac{1}{2} \|V - WH\|_F^2 + \lambda \sum_{j=1}^r \|H(j, :)\|_1, W, H \geq 0 \quad (1)$$

Here, $\|V - WH\|_F^2$ is the non-negative matrix factorization used for stain separation and $\|H(j, :)\|_1$ is L1 sparseness regularization on stains H . V is the relative optical density, W is an $m \times r$ matrix called the stain color matrix. The columns of W include the RGB color of each stain (in our case, $m=3$ RGB channels and $r =$ number of stains). H is an $r \times n$ matrix called stain concentration matrix ($n =$ number of pixels). The rows of H represent the total amount of stained tissue. λ is a sparsity regularization parameter. The objective function of the stain normalization method is solved by exchanging between W and H (i.e., optimizing one of them while fixing the other).

3.2 Individual cell nuclei segmentation models (ISs)

In the literatures, there are several deep CNN-based semantic segmentation methods that achieved promising results with various image modalities. We assessed the performance of several efficient deep CNN-based semantic segmentation models with the cell nuclei segmentation problem [27]. Based on the analysis [27], we selected the three best cell nuclei segmentation models ($N=3$), namely the FCDenseNet, SegNet and Self-Correction models to develop the individual nuclei segmentation models (ISs).

As shown in the studies [11, 28], FCN obtained good segmentation results with natural images using simple CNN architectures. FCN can handle the images regardless of their size. Several improved networks have been proposed in the literature on the basis of FCN to further improve the segmentation result.

FCDenseNet [29] is an extension of the densely connected convolutional networks (DenseNets) [30] proposed to handle the semantic segmentation problem. The key idea [29] is to connect each layer to all other layers following a feed-forward approach. Let x_n be the feature maps of the n^{th} layer of a standard CNN, a non-linear transformation H_n is applied to the feature maps of the previous layer x_{n-1} to calculate x_n as follows:

$$x_n = H_n(x_{n-1}) \quad (2)$$

The non-linear transformation H is comprises a convolution layer followed by a rectified linear activation function ($ReLU(x) = \max(x, 0)$) and a dropout [30]. As shown in Figure 2(a), all previous feature maps are concatenated to construct the input of each DenseBlock. This process can be formulated as follows:

$$x_n = Q_n([x_{n-1}, x_{n-2}, x_{n-3}, \dots, x_0]) \quad (3)$$

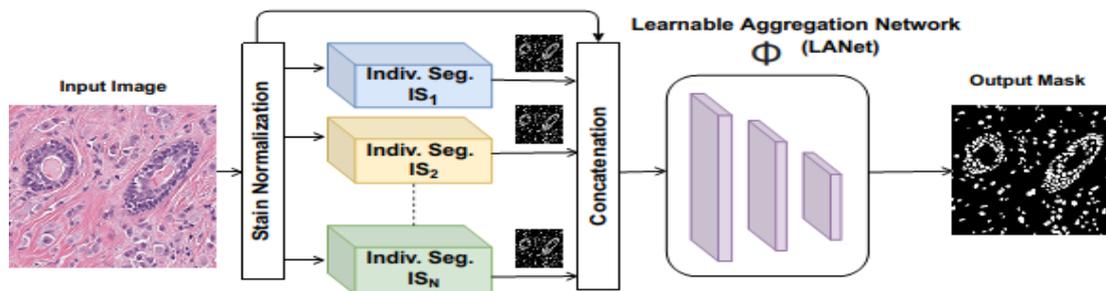


Figure 1. Overview of the proposed framework for nuclei segmentation

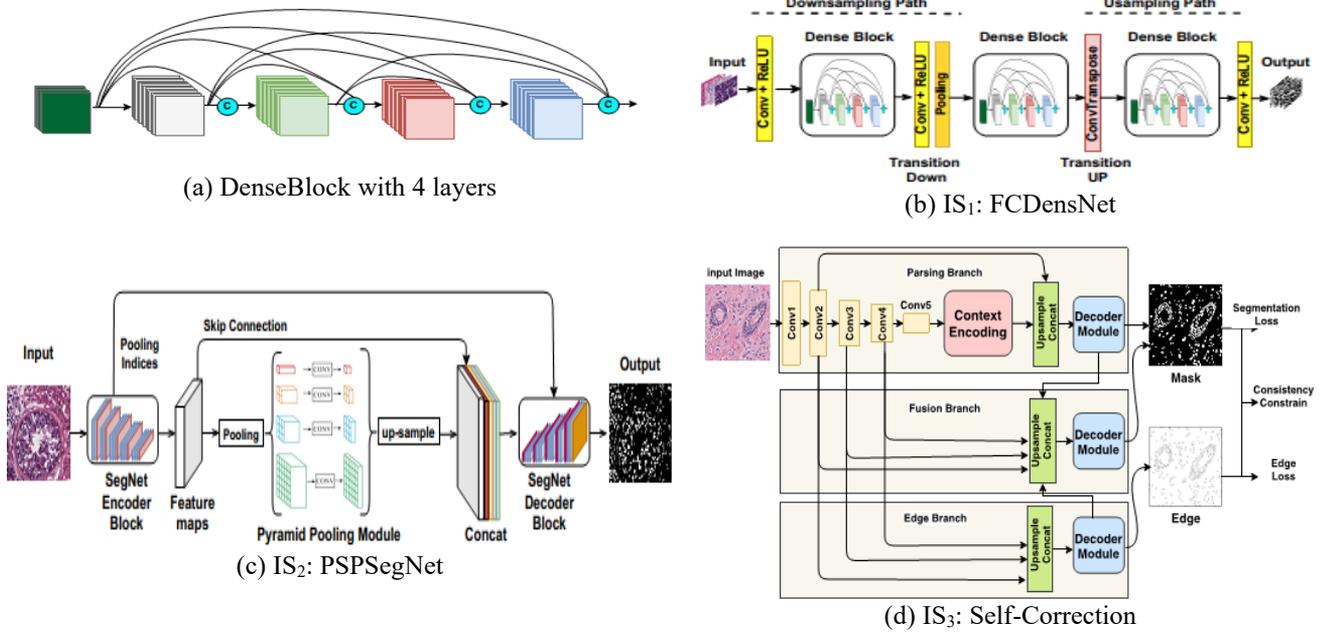


Figure 2. Architectures of the individual nuclei segmentation models used in this study. (a) DenseBlock, (b) FCDenseNet (IS₁), (c) PSPSegNet (IS₂) and (d) Self-Correction (IS₃)

In this expression, $[.]$ stands for the feature maps concatenation. In this case, the non-linear transformation Q includes a batch normalization (BN), ReLU, convolution and dropout layers [30]. The main aim of DenseNets is to reuse feature maps and prevent the feature explosion problem at the layers of the up-sampling path. FCDensNet comprises a down-sampling path, up-sampling path and skip connections (it follows the architecture of FCN). For simplicity, Figure 2(b) shows FCDenseNet with one denseblock in each of down-sampling path, up-sampling path and in between as bottleneck. In this study, we adopt the FCDenseNet103 architecture described in Ref. [24]. FCDenseNet103 includes 103 convolutional layers. The input layer that receives the WSI images, the down-sampling path has 38 layers, the bottleneck has 15 layers while the up-sampling path has 38 layers. In our study, the trained FCDenseNet103 model includes five transition-down (TD) and five transition-up (TU). Each TU contains a transposed convolution layer. In turn, each TD comprises convolution, ReLU and Pooling layers. It should be noted that FCDenseNet103 classifies each pixel in the input WSI image should be classified as cell nuclei or non-cell nuclei pixel. Thus, the top layer in the FCDenseNet103 network includes a 1×1 convolution and softmax non-linearity layers to produce the per class distribution of each pixel in the input WSI image.

PSPSegNet architecture is constructed by combining the pyramid scene parsing network (PSPNet) [31] with the SegNet [13] model, as shown in Figure 2(c). The SegNet architecture includes encoder layers, decoder layers and a pixel-wise classification layer. The first 13 convolutional layers of VGG16 [32] (VGG16 without the fully connected layers) are used to form the encoder of SegNet. In the decoder network, the max-pooling indices received from the corresponding encoder layers are used to up-sample the feature maps. Indeed, the use of the max-pooling indices eliminates the need for up-sample learning techniques. PSPNet model has been designed to use the global context information of the cell nuclei in the WSI images by employing a deep residual network (ResNet) [5] to extract different patterns from the input images. To

extract patterns of different scales, the feature maps are fed into a pyramid pooling module. The resulting multi-scale feature maps are pooled and fed into a 1×1 convolutional layer to reduce the size of the extracted features. To capture the local and global context information of the cell nuclei in the WSI images, the feature maps of the pyramid pooling module are up-sampled and concatenated with the inputted feature maps. Finally, a convolutional layer is used at the top of the network to produce the pixel-wise predictions.

It should be noted that we replaced the VGG16 encoder network with ResNet101 in the SegNet network. We used the pyramid pooling module amidst encoder and decoder networks. Besides, we concatenated the feature maps of the encoder with the up-sampled feature maps of the pyramid pooling module. Then, we fed the concatenated feature maps into the decoder of the SegNet model.

The self-correction model proposed by Li et al. [33] uses a self-correction training strategy for developing a human parsing (SCHP) model. The network architecture of SCHP is influenced by the CE2P network proposed by Ruan et al. [34]. As shown in Figure 2(d), the self-correction model includes three streams: parsing, edge, and fusion. The loss function of A-CE2P can be expressed as follows:

$$E = \alpha_1 E_{segmentation} + \alpha_2 E_{consistent} + \alpha_3 E_{edge} \quad (4)$$

There are three different losses: $E_{segmentation}$, $E_{consistent}$, and E_{edge} , each corresponds to a stream. A-CE2P uses three weights (α_1 , α_2 and α_3) to control the contribution of each loss. It should be noted that the A-CE2P model is jointly trained by optimizing the loss function E in an end-to-end manner. After obtaining acceptable nuclei segmentation masks using the A-CE2P segmentation model, a self-correction mechanism uses a cyclically learning scheduler with warm restarts [35]. A set of weights (models), $W = \{\hat{w}_0, \hat{w}_1, \dots, \hat{w}_N\}$ and the corresponding predicted labels, $Y = \{\hat{y}_0, \hat{y}_1, \dots, \hat{y}_N\}$ are produced after each cycle of the self-correction mechanism. The weights \hat{w} of the current cycle of

the self-correction mechanism are combined with the weights of the preceding cycle \hat{w}_{n-1} to produce updated weights \hat{w}_n as follows:

$$\hat{w} = \frac{n}{n+1} \hat{w}_{n-1} + \frac{1}{n+1} \hat{w} \quad (5)$$

Similarly, the predicted labels are aggregated as follows:

$$\hat{y} = \frac{n}{n+1} \hat{y}_{n-1} + \frac{1}{n+1} \hat{y} \quad (6)$$

Here, n donates the current cycle number and $0 \leq n \leq N$, and \hat{y} stands for the pseudo-labels (i.e., pseudo masks) generated by the model \hat{w}_n .

3.3 Learnable Aggregation Network (LANet)

Let M_i be a coarse nuclei mask generated by IS_i , and M_0 is the input WSI image. The input I for LANet is the concatenation of the coarse masks and the input WSI image:

$$I = \bigcup_{i=0 \rightarrow n} M_i \quad (7)$$

In this study, n is set to 3 (three individual models) and $i = 0$ stands for the input WSI image (i.e., M_0). It should be noted that in this case, we have an input WSI image with size of $W \times H \times 3$ (width \times height \times number of channels) along with the three coarse masks (each mask has a size of $W \times H \times 1$) which are contiguous along the color channel axis. We stacked the WSI image and the three masks together along the third dimension (i.e., the channel dimension), and thus the input to the LANet, I , has a size of $W \times H \times 6$.

Generally speaking, an aggregation operator Φ could be represented by a mathematical formula to combine a set of values into a single value. In our case, aggregation operators are employed to combine N coarse masks generated by ISs into a single mask having fine labels. In this context, the weighted average is the most well-known aggregation operator, however, it is difficult to compute optimal and generalized weights. Therefore, our intuition here is to develop a CNN-based aggregation network (i.e., LANet network), in which the weights are optimally computed through the back-propagation process. The aggregation function based on LANet can be expressed as follows:

$$F = \phi(M_0, M_1, \dots, M_n) = \sigma(BN(\sum_i w_i x_i + b)) \quad (8)$$

Here, σ is a non-linear activation ReLU function, BN is the batch normalization and w_i and b are the weights of LANet. Table 1 presents the architecture details of LANet used on the proposed nuclei segmentation method as aggregation model Φ .

Table 1. Architecture of LANet

Architecture
Input, ch = 6
ConvBlock (2 layers), ch = 32
ConvBlock (2 layers), ch = 64
ConvBlock (2 layers), ch = 128
UpConv, ch = 64
UpConv, ch = 32
UpConv, ch = 2

Table 2. Building blocks of LANet

ConvBlock	UpConv
3 \times 3 Convolution	Up Sample (factor =2)
Batch Normalization	3 \times 3 Convolution
3 \times 3 Convolution	Batch Normalization
Batch Normalization	ReLU
ReLU	
2 \times 2 MaxPooling	

As one can see in Table 2, the architecture of LANet follows the UNet. LANet comprises three convolution blocks in the down-sampling path and three up-sampling blocks in the up-sampling path. Each convolution block includes two 3 \times 3 convolution layers, batch normalization, ReLU and 2 \times 2 MaxPooling.

In turn, each block in the up-sampling path includes an up-sampling layer (up-sampling factor=2), 3 \times 3 convolution layers, batch normalization and ReLU. It is worth noting that the concatenation of the input WSI image which contains the low-level row information with the coarse nuclei segmentation masks which highlights the important regions (Eq. (7)) can significantly recuperate missed segmented cells by IS models through guiding LANet network and helping to refine final mask.

3.4 Implementation of the cell nuclei segmentation method

Algorithm 1 presents the implementation steps of the proposed cell nuclei segmentation method for multiple organs and multiple centers WSI images.

Algorithm 1. Implementation of the proposed method

1:	Read WSI images.
2:	Perform stain normalization for all WSI images.
3:	Split the WSI dataset into train dataset and test sets.
4:	for $i = 1: N$ do
5:	Train IS_i using train set.
6:	Save trained IS_i model.
7:	end
8:	Obtain coarse masks from each IS.
9:	Concatenate N masks with the WSI image, Eq. (7).
10:	Train the aggregation model Φ
11:	Save Aggregation model Φ
12:	foreach image in test dataset do
13:	Read test image
14:	for $i = 1: N$ do
15:	Get coarse mask of IS_i
16:	end
17:	Concatenate N masks with the WSI images, Eq. (7).
18:	Get output mask from aggregation model Φ
19:	end

The key steps of the training and test phases of the proposed method are highlighted below:

- The dataset is split into training and testing sets. The stain of all WSI is normalized. Then, N accurate ISs are developed for nuclei segmentation using the training set.
- The coarse masks of ISs are then concatenated along with the input WSI image, generating a new feature map with $n + 3$ channels input.
- The generated feature map is then fed into LANet to produce an accurate segmentation mask.

To assess the performance of the proposed cell nuclei segmentation method, we use two widely used evaluation metrics for cell nuclei image segmentation, namely the

aggregated Jaccard index (AJI) [18] and F1-score (F1). However, AJI is a modified version of the Jaccard index, it has a more powerful ability to measure the strength of segmentation. AJI can be computed as follows:

$$AJI = \frac{\sum_{i=1}^L |GT_i \cap NP_j^*(i)|}{\sum_{i=1}^K |GT_i \cup NP_j^*(i)| + \sum_{K \in Ind} NP_k} \quad (9)$$

Here, GT_i is the i th ground-truth mask of nuclei pixels, NP_i is the predicted nuclei segmentation mask, $NP_j^*(i)$ stands for the connected component of the predicted cell nuclei segmentation mask that maximizes the Jaccard index, and Ind includes the indices of pixels that do not belong to the ground-truth.

The F1-score can be expressed as follows:

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \quad (10)$$

$$precision = \frac{TP}{TP + FP} \quad (11)$$

$$recall = \frac{TP}{TP + FN} \quad (12)$$

In these equations, true positive (TP) stands for the number of correctly segmented cell nuclei pixels. True negative (TN) stands for the number of correctly segmented non-cell nuclei pixels. False positive (FP) stands for the number of non-cell nuclei pixels wrongly segmented as nuclei. False negative (FN) stands for the number of cell nuclei pixels wrongly segmented as non-cell nuclei. The F1-score is the harmonic mean of the precision and recall.

4. EXPERIMENTS AND RESULTS

4.1 Dataset

In our experiments, we use a challenging multiple organ and multiple center dataset called MoNuSeg dataset [18] to validate the proposed cell nuclei segmentation method. This dataset has been prepared by the Indian Institute of Technology Guwahati. MoNuSeg dataset includes annotated H&E stained WSI images captured at 40x magnification. The dataset comprises a total of 30 WSI images with around 22,000 nuclear boundary annotations from colon, breast, stomach, kidney, prostate, bladder and liver organs. The WSI images were collected at different medical centers (image size is 1000 × 1000).

4.2 Implementation and training details

In our experiments, we split the MoNuSeg dataset as follows: 23 WSI images for training and 7 WSI images for testing (the test set includes one WSI image of each organ). The resolution of the WSI images is 1000 × 1000 pixels. The input patch size of the IS models varies from 512 × 512 to 256 × 256. Specifically, to train the ISs we scaled each image in the training set (23 WSI images) to 1024 × 1024. Four non-overlapping sub-images of size 512 × 512 were extracted from each scaled image. Besides, we extracted a total of 200

randomly cropped patches with a size of 512 × 512 from each image in the training test. In total, 4692 patches (23 × 4 + 23 × 200) of size 512 × 512 were used to train the ISs. Note that we could avoid overfitting with this amount of data. It should be noted that the variation in the input image size is due to the design restrictions of each architecture. Here, the input image sizes of FCDenseNet, PSPSegNet, and self-Correction are 512 × 512, 256 × 256, and 128 × 128, respectively. To train LANet, a total of 2300 non-overlapped sub-images of size 128 × 128 were extracted from the coarse masks of the IS models. In turn, we scaled each testing image to 1024 × 1024 and then we divided it into four non-overlapping sub-images of size 512 × 512. Of note, the resulting cell nuclei masks corresponding to the four patches of each testing image were stacked while keeping the spatial order to get the final cell nuclei mask.

It should be noted that the batch size used in our experiments was 2 and the number of epochs was 100. We used the cross-entropy loss function and the stochastic gradient descent (SGD) optimizer for training all segmentation models. The values of the initial learning rate, momentum and weight decay were 1e-1, 0.99 and 1e-8, respectively. All models were implemented with python3.8 and pytorch1.5. The experiments were carried out in a computer with the following specification: AMD Ryzen 5 3600 6-core CPU, 32 GB memory and Nvidia RTX2060 8GB GPU.

4.3 Results and discussion

Table 3 presents the F1 and AJI-scores of the IS models as well as the proposed method. As one can see, the three developed individual nuclei segmentation models obtain promising segmentation results (AJI Score >69% and F1 score >87%). The proposed method achieves a noticeable improvement on both F1-score and AJI-score.

Table 3. Evaluation of the proposed method

Model	F1-Score	AJI-Score
FCDenseNet103	88.45%	71.34%
PSPSegNet	88.42%	70.95%
Self-Correction	87.45%	69.23%
LANet w/o concat. RGB	89.21%	72.59%
Proposed	89.25%	73.02%

Our method obtains an F1 and AJI scores of 89.21% and 72.59%, respectively, without concatenating the input WSI with the coarse nuclei masks (LANet w/o concat. RGB). However, concatenating the WSI with the coarse nuclei masks at the aggregation network improves the F1 score to 89.25% and the AJI Score to 73.02%. The obtained cell nuclei segmentation results prove that the suggested aggregation approach, LANet, not only can aggregate the coarse masks of the IS models, but also it can refine the coarse labels and produce accurate nuclei segmentation masks.

Figures 3 and 4 show the boxplots of the F1 and AJI scores of the ISs and our method. As shown, the proposed cell nuclei segmentation method achieves the highest median AJI and F1 scores (median F1 score 85% and median AJI 73% approximately.) when compared to the IS models. It is worth noting that all cell nuclei segmentation methods have not outliers on both evaluation metrics and the minimum values of F1 and AJI scores are higher than 78% and 64% respectively. This analysis confirms the suitability of the IS models as candidates for constructing the proposed nuclei segmentation method.

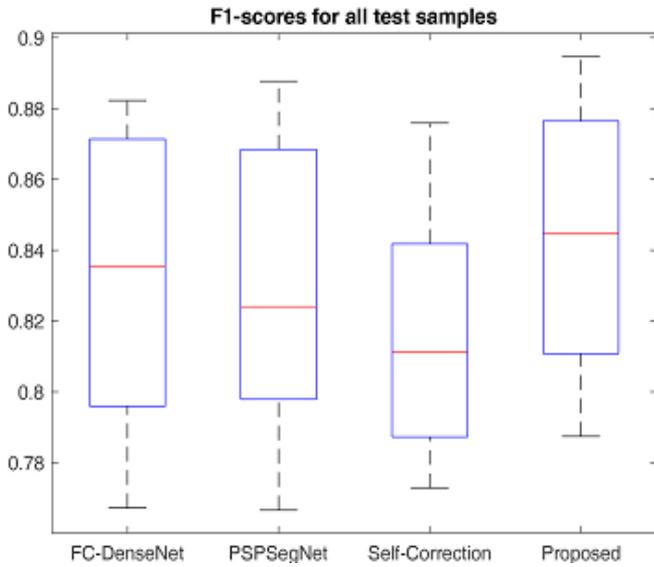


Figure 3. Boxplots of the F1-score for the IS models and the proposed method

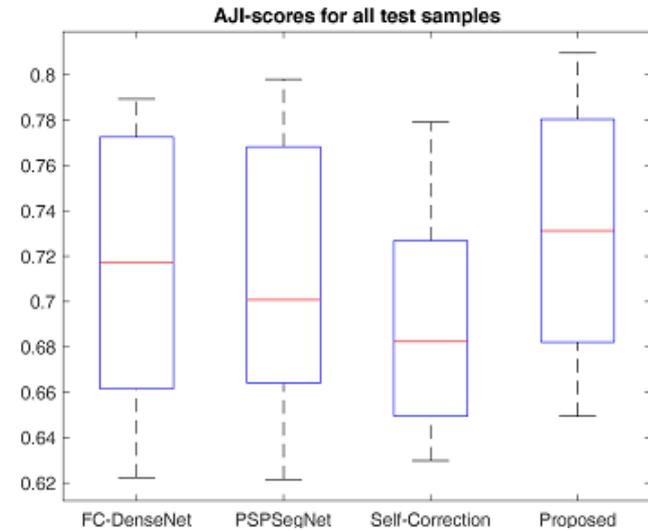


Figure 4. Boxplots of the AJI-score for the IS models and proposed method

Figure 5 compares the cell nuclei segmentation masks of the individual nuclei segmentation models and the proposed method for WSI images of different organs. It is worth noting that we compare our method with three recently published cell nuclei segmentation methods: FCDenseNet [27], SegNet [27] and Self-Correction [35]. As one can see, the proposed method produces accurate segmentation results for the different organs. For instance, the proposed method can accurately segment the WSI images of the liver and bladder organs that have big cell nuclei with intense overlapping. Specifically, the proposed method achieves AJI and F1 scores of 72 and 81% respectively, with these WSI images, noting that the AJI score of the proposed method is 1.25 points higher than the IS models. Besides, in the case of the WSI images of the breast, colon and prostate organs, the proposed method surpasses the IS models with AJI scores of 67, 65 and 74% respectively (approximately 1.1 points higher). Also, Figure 5 presents the WSI images of kidney and stomach organs that have a dense number of cell nuclei. With these images, the proposed method achieves AJI scores of 73 and 80% respectively (an improvement of 1.1 points when compared with the IS models).

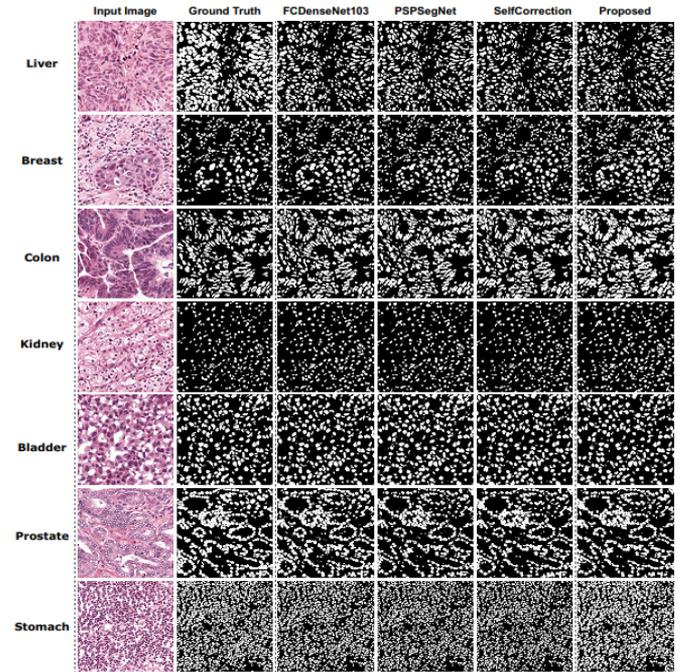


Figure 5. Segmentation results of the proposed cell nuclei segmentation method, IS₁--FC-DenseNet [27], IS₂--PSPSegNet [27] and IS₃--Self-Correction [35] (three recently published nuclei segmentation models) with the WSI images of the liver, breast, colon, kidney, bladder, prostate, and stomach organs

Table 4 compares the cell nuclei segmentation results of our method and some state-of-the-art semantic segmentation networks (UNet [8], FCN8s [11] and SegNet [13]) trained to segment cell nuclei in WSI images. Also, the proposed method is compared with recently published cell nuclei segmentation networks ([10, 12, 14, 15]) along with a popular segmentation software [16]. All models have been trained on the MoNuSeg dataset [18] under the same experimental conditions. As one can see in Table 4, our method beats the compared cell nuclei segmentation methods. Specifically, the proposed achieves AJI score 1.10 points higher than the UNet [8] and SegNet [13]. Also, it achieves an F1 score 1.30 points higher than [14, 15]. It should be noted that the cGANs model [10] has been trained with original WSI and synthesized WSI images. However, the cGANs model [10] is a heavy model with 58 million approximately trainable. It obtains 1% AJI score lower than our method. Additionally, we compared our method with a popular segmentation software called ImageFiji [16]. ImageFiji obtains an AJI score much lower than the proposed method.

Table 4. Comparing the F1 and AJI scores of the proposed cell nuclei segmentation method with state-of-the-art eight methods

Model	F1-Score	AJI-Score
FCN8s [11]	88.17%	70.86%
UNet [8]	87.84%	70.02%
SegNet [13]	86.48%	68.28%
UNet++ [12]	88.24%	70.88%
RIC-UNet [14]	82.78%	56.35%
DIST [15]	78.63%	55.98%
ImageFiji [16]	75.52%	53.30%
cGANs [10]	86.60%	72.10%
Proposed	89.25%	73.02%

Based on the above analysis, the proposed cell nuclei segmentation method with a graphical user interface (GUI) could be a powerful fully automated tool that would highly assist pathologists in obtaining fast and accurate cell nuclei delineation for WSI images of different organs collected from different centers.

5. CONCLUSIONS

This article has proposed an accurate cell nuclei segmentation method based on a deep learnable aggregation network called LANet and a set of accurate deep CNN-based individual cell nuclei segmentation models. With a challenging multiple organ and multiple center WSI image dataset, the proposed method achieved an AJI score of 73%, which is 3 points higher than UNet (the popular medical image segmentation model), and 1 point higher than the cGAN-based cell nuclei segmentation model [10]. Future research will be focused on the use of different aggregation approaches to fuse robust deep learning-based cell nuclei segmentation models to further enhance the segmentation accuracy and reliability.

ACKNOWLEDGMENT

This research was partially supported by the Government of Spain through Project under Grant PID2019-105789RB-I00.

REFERENCES

[1] Madabhushi, A., Lee, G. (2016). Image analysis and machine learning in digital pathology: Challenges and opportunities. *Medical Image Analysis*, 33: 170-175. <https://doi.org/10.1016/j.media.2016.06.037>

[2] Lu, C., Romo-Bucheli, D., Wang, X., Janowczyk, A., Ganesan, S., Gilmore, H., Rimm, D., Madabhushi, A. (2018). Nuclear shape and orientation features from H&E images predict survival in early-stage estrogen receptor-positive breast cancers. *Laboratory Investigation*, 98(11): 1438-1448. <https://doi.org/10.1038/s41374-018-0095-7>

[3] Krizhevsky, A., Sutskever, I., Hinton, G.E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6): 84-90. <https://doi.org/10.1145/3065386386>

[4] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q. (2017). Densely connected convolutional networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2261-2269. <https://doi.org/10.1109/cvpr.2017.243>

[5] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778. <https://doi.org/10.1109/cvpr.2016.90>

[6] Irshad, H., Veillard, A., Roux, L., Racoceanu, D. (2014). Methods for nuclei detection, segmentation, and classification in digital histopathology: A review—Current status and future potential. *IEEE Reviews in Biomedical Engineering*, 7: 97-114. <https://doi.org/10.1109/rbme.2013.2295804>

[7] Ciresan, D., Giusti, A., Gambardella, L., Schmidhuber, J. (2012). Deep neural networks segment neuronal

membranes in electron microscopy images. *Advances in Neural Information Processing Systems*, 25: 2843-2851.

[8] 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*, pp. 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

[9] Arai, K., Kapoor, S. (2020). Advances in computer vision. *Proceedings of the 2019 Computer Vision Conference (CVC)*. <https://doi.org/10.1007/978-3-030-17798-0>

[10] Mahmood, F., Borders, D., Chen, R.J., McKay, G.N., Salimian, K.J., Baras, A., Durr, N.J. (2020). Deep adversarial training for multi-organ nuclei segmentation in histopathology images. *IEEE Transactions on Medical Imaging*, 39(11): 3257-3267. <https://doi.org/10.1109/tmi.2019.2927182>

[11] 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*, pp. 3431-3440. <https://doi.org/10.1109/cvpr.2015.7298965>

[12] Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J. (2020). UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging*, 39(6): 1856-1867. <https://doi.org/10.1109/tmi.2019.2959609>

[13] Badrinarayanan, V., Kendall, A., Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12): 2481-2495. <https://doi.org/10.1109/tpami.2016.2644615>

[14] Zeng, Z., Xie, W., Zhang, Y., Lu, Y. (2019). RIC-Unet: An improved neural network based on Unet for nuclei segmentation in histology images. *IEEE Access*, 7: 21420-21428. <https://doi.org/10.1109/access.2019.2896920>

[15] Naylor, P., Lae, M., Reyal, F., Walter, T. (2019). Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE Transactions on Medical Imaging*, 38(2): 448-459. <https://doi.org/10.1109/tmi.2018.2865709>

[16] Dong, F., Irshad, H., Oh, E.Y., Lerwill, M.F., Brachtel, E.F., Jones, N.C., Knoblauch, N.W., Montaser-Kouhsari, L., Johnson, N.B., Rao, L.K.F., Faulkner-Jones, B., Wilbur, D.C., Schnitt, S.J., Beck, A.H. (2014). Computational pathology to discriminate benign from malignant intraductal proliferations of the breast. *PLoS One*, 9(12): e114885. <https://doi.org/10.1371/journal.pone.0114885>

[17] Pan, X., Li, L., Yang, D., He, Y., Liu, Z., Yang, H. (2019). An accurate nuclei segmentation algorithm in pathological image based on deep semantic network. *IEEE Access*, 7: 110674-110686. <https://doi.org/10.1109/access.2019.2934486>

[18] Kumar, N., Verma, R., Sharma, S., Bhargava, S., Vahadane, A., Sethi, A. (2017). A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE Transactions on Medical Imaging*, 36(7): 1550-1560. <https://doi.org/10.1109/tmi.2017.2677499>

[19] Naylor, P., Laé, M., Reyal, F., Walter, T. (2017). Nuclei segmentation in histopathology images using deep neural

- networks. In 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), pp. 933-936. <https://doi.org/10.1109/isbi.2017.7950669>
- [20] Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A. (2017). Inception-v4, inception-ResNet and the impact of residual connections on learning. In Proceedings of the AAAI Conference on Artificial Intelligence.
- [21] Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E. (2020). Squeeze-and-excitation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(8): 2011-2023. <https://doi.org/10.1109/tpami.2019.2913372>
- [22] Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O. (2006). CellProfiler: Image analysis software for identifying and quantifying cell phenotypes. *Genome Biology*, 7(10): 1-11. <https://doi.org/10.1186/gb-2006-7-10-r100>
- [23] Graham, S., Vu, Q.D., Raza, S.E.A., Azam, A., Tsang, Y.W., Kwak, J.T., Rajpoot, N. (2019). Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Analysis*, 58: 101563. <https://doi.org/10.1016/j.media.2019.101563>
- [24] Qu, H., Wu, P., Huang, Q., Yi, J., Yan, Z., Li, K., Riedlinger, G.M., De, S., Zhang, S., Metaxas, D.N. (2020). Weakly supervised deep nuclei segmentation using partial points annotation in histopathology images. *IEEE Transactions on Medical Imaging*, 39(11): 3655-3666. <https://doi.org/10.1109/tmi.2020.3002244>
- [25] Graham, S., Rajpoot, N.M. (2018). SAMS-NET: Stain-aware multi-scale network for instance-based nuclei segmentation in histology images. In 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 590-594. <https://doi.org/10.1109/isbi.2018.8363645>
- [26] Vahadane, A., Peng, T., Albarqouni, S., Baust, M., Steiger, K., Schlitter, A.M. (2015). Structure-preserved color normalization for histological images. In 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), pp. 1012-1015. <https://doi.org/10.1109/isbi.2015.7164042>
- [27] Hassan, L., Saleh, A., Abdel-Nasser, M., Omer, O.A., Puig, D. (2021). Promising deep semantic nuclei segmentation models for multi-institutional histopathology images of different organs. *International Journal of Interactive Multimedia and Artificial Intelligence*, 6(6): 35. <https://doi.org/10.9781/ijimai.2020.10.004>
- [28] Ulku, I., Akagunduz, E. (2019). A survey on deep learning-based architectures for semantic segmentation on 2d images. arXiv preprint arXiv:1912.10230.
- [29] Jégou, S., Drozdal, M., Vazquez, D., Romero, A., Bengio, Y. (2017). The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 11-19. <https://doi.org/10.1109/cvprw.2017.156>
- [30] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700-4708. <https://doi.org/10.1109/cvpr.2017.243>
- [31] Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J. (2017). Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2881-2890. <https://doi.org/10.1109/cvpr.2017.660>
- [32] Simonyan, K., Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
- [33] Li, P., Xu, Y., Wei, Y., Yang, Y. (2021). Self-correction for human parsing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1-1. <https://doi.org/10.1109/tpami.2020.3048039>
- [34] Ruan, T., Liu, T., Huang, Z., Wei, Y., Wei, S., Zhao, Y. (2019). Devil in the details: Towards accurate single and multiple human parsing. In Proceedings of the AAAI Conference on Artificial Intelligence, 33(1): 4814-4821. <https://doi.org/10.1609/aaai.v33i01.33014814>
- [35] Abdel-Nasser, M., Saleh, A., Puig, D. (2020). Channel-wise aggregation with self-correction mechanism for multi-center multi-organ nuclei segmentation in whole slide imaging. In VISIGRAPP, 4: 466-473. <https://doi.org/10.5220/0009156604660473>