Promising Deep Semantic Nuclei Segmentation Models for Multi-Institutional Histopathology Images of Different Organs

Loay Hassan[1,3*], Adel Saleh[2], Mohamed Abdel-Nasser[1,4], Osama A. Omer[1], Domenec Puig[4]

[1] Electrical Engineering Department, Aswan University (Egypt)
[2] Gaist Solutions Ltd, Skipton BD23 2TZ (UK)
[3] Department of computer science, Arab Academy for Science, Technology and Maritime Transport, Aswan (Egypt)
[4] Department of Computer Engineering and Mathematics, University Rovira i Virgili, 43007 Tarragona (Spain)

unir
LA UNIVERSIDAD
EN INTERNET

Abstract

Nuclei segmentation in whole-slide imaging (WSI) plays a crucial role in the field of computational pathology. It is a fundamental task for different applications, such as cancer cell type classification, cancer grading, and cancer subtype classification. However, existing nuclei segmentation methods face many challenges, such as color variation in histopathological images, the overlapping and clumped nuclei, and the ambiguous boundary between different cell nuclei, that limit their performance. In this paper, we present promising deep semantic nuclei segmentation models for multi-institutional WSI images (i.e., collected from different scanners) of different organs. Specifically, we study the performance of pertinent deep learning-based models with nuclei segmentation in WSI images of different stains and various organs. We also propose a feasible deep learning nuclei segmentation model formed by combining robust deep learning architectures. A comprehensive comparative study with existing software and related methods in terms of different evaluation metrics and the number of parameters of each model, emphasizes the efficacy of the proposed nuclei segmentation models.

Keywords

I. Introduction

Nowadays, digital pathology is rapidly gaining momentum as a proven and essential technology. Its popularity has grown in the last decade due to the improvements in hardware and software. The whole-slide imaging (WSI) refers to the scanning of conventional glass slides to produce high-resolution digital images slides, that can be stored and accessed using dedicated software. The potential applications of digital pathology comprise cell segmentation, counting cancer cells, and prognosis of cancers.

Cell segmentation refers to the process of identifying groups of pixels that represent cell nuclei. This process is often complicated, especially in the presence of adjacent or overlapping cells and color variation in histopathological images. It is one of the core operations in histopathology image analysis. So, in the context of computational pathology, accurate nuclei segmentation techniques are highly needful for extracting, mining, and interpreting sub-cellular morphologic information from digital slide images. Several extracted descriptors such as cell nuclei shape and number of cell nuclei in WSI images are key components of studies such as the determination of cancer types, cancer grading, and prognosis [1].

Indeed, there is diverse tissue types, variations in staining, and cell types, leading to different visual characteristics of WSI images. These variations make the segmentation of nuclei segmentation a challenging task (see Fig. 1). The visual characteristics of WSI images make it very difficult to develop traditional image processing-based segmentation algorithms that give acceptable nuclei segmentation results. The difficulty increases when the segmentation algorithms handle WSI images taken from several cancer patients and collected at different medical centers for various organs, such as breast, kidney, prostate, and stomach [4]. Existing nuclei segmentation software and toolboxes include Cell profiler [2] and ImageJ-Fiji [3]. Cell Profiler simultaneously measures the size, shape, intensity, and texture of a variety of cell types in a high throughput manner [2]. ImageJ-Fiji exploits the latest software engineering practices to merge powerful software libraries with a wide range of scripting languages to allow fast prototyping of image processing algorithms [3].

In turn, the success of deep learning models with several computer vision-based applications encouraged researchers to make extensive efforts of works attempted at developing image segmentation approaches using deep learning models. Of note, the nuclei segmentation task necessitates an enormous effort to manually create

pixel-wise annotations to be used for training deep learning models. For instance, a multi-path dilated residual network was proposed in [14] for nuclei segmentation and detection. In [17], a nuclei segmentation method based on deep convolutional neural networks (DCNNs) for histopathology images was proposed. However, existing nuclei segmentation methods may achieve good results with a dataset of WSI images and poor performance with other datasets. The main reasons for these limited results are color variations in histopathological images resulted from acquiring WSI from different scanners, the overlapping and clumped nuclei, and the ambiguous boundary between adjacent cell nuclei.
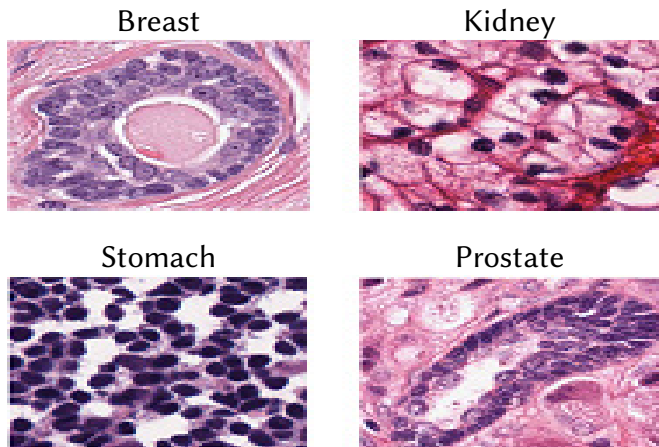

Fig. 1. Examples of WSI images for multi-organ samples.

To tackle these challenges, in this paper, we present promising deep semantic nuclei segmentation models for multi-institutional histopathology images of different organs. These deep learning-based semantic segmentation models are trained in a supervised way to focus on the nuclear regions and to discriminate between nuclear pixels and other pixels. In this way, the models can learn nuclei-aware features, color information, as well as recognizing the complete cells. Indeed, such promising nuclei segmentation models can be used to extract apt features for nuclear morphometrics. Also, it could contribute to the advancement of digital pathology software.

The key contributions of this paper are:

- Study the performance of different deep learning models with nuclei segmentation in WSI images of various stains and various organs. A challenging multi-institutional multi-organ WSI image dataset is used in this paper (publicly available dataset).

- Propose a feasible deep learning nuclei segmentation model formed by combining robust deep learning architectures (so-called PSPSegNet). It achieves 3.48% improvement on the F1-score and 6.62% improvement on aggregated Jaccard index (AJI).

- A comprehensive comparative study with existing software and related methods is presented, in terms of different evaluation metrics and the number of parameters of each model. Also, the use of nuclei segmentation models to count the number of nuclei in WSI images.

Below, we present the remaining sections of this paper. In Section II, we study and discuss the related work. In Section III, we explain the methodology in detail. Section IV includes the experimental results, comparisons and discussion. Section V concludes the paper and gives different points of future work.

II. Related Work

In the last years, various deep learning models have been employed for performing different segmentation tasks in biology [4]. Generally, most outstanding deep segmentation models are based on convolutional neural networks (CNNs), recurrent neural networks (RNNs), encoder-decoders architecture, and generative adversarial networks (GANs).

Naylor et al. [5] introduced a fully automated method for cell nuclei segmenting in WSI images based on three segmentation models, namely PangNet, a fully convolutional network (FCN), and DeconvNet. They ensembled the three segmentation models, obtaining an F1-score of 0.80. Also, Naylor et al. [6] proposed a segmentation method of nuclei in histopathology data based on CNN. They proposed a distinct idea to segment toughing or overlapping nuclei by formulating the problem as a regression task, where they aim at predicting the distance map of nuclei. They claimed that the main problem in the segmentation of nuclei is that segmentation methods tend to segment adjacent or overlap nuclei one object. Their approach outperforms some related nuclei segmentation methods on the AJI score.

Wang et al. [7] proposed a bending loss regularized network for nuclei segmentation in histopathology images. The proposed bending loss defines high penalties to contour points with large curvatures and applies small penalties to contour points with a slight curvature. Minimizing bending loss can avoid generating contours that encompass multiple nuclei. In the case of histopathology images, nuclei have a smooth shape, and the points on the boundaries of nuclei have small curvature changes. In turn, the points on the contour with large curvature changes have a high probability of being the touching points of two or multiple nuclei. The nuclei segmentation scheme comprises three steps: 1) a preprocessing step for color normalization, 2) an encoder-decoder architecture with the bending loss, and 3) a postprocessing step described in [8] was employed. The proposed model was validated on the MoNuSeg dataset, obtaining an AJI score of 0.621 with the same organ test and score of 0.641 with different organ tests.

Al-Kofahi et al. [9] proposed a three-step cell nuclei segmentation approach: 1) the detection of the cells using a deep learning-based model to obtain pixel probabilities for nuclei, cytoplasm, as well as background, 2) the separation of touching cells based on blob detection and shape-based watershed techniques that can distinguish between the individual nuclei from the nucleus prediction map, and 3) the segmentation of the nucleus and cytoplasm. With four different datasets, they obtained an accuracy of 0.84. Besides, Cui et al. [10] proposed an automatic end-to-end deep neural network algorithm for the segmentation of individual nuclei. They introduced a nucleus-boundary model to predict nuclei and their boundaries simultaneously using a fully convolutional neural network. They obtained the area of each nucleus via a simple, fast, and parameter-free postprocessing procedure. This method can segment a *1000x1000* image in less than 5 seconds, which facilitates precisely segment WSI images in an acceptable time.

In [11], Qu et al. proposed a weakly supervised segmentation framework based on partial points annotation in histopathology images. The framework consists of two stages: 1) a semi-supervised strategy to learn a detection model, and 2) a segmentation model is trained from the detected nuclei locations in a weakly-supervised manner. Specifically, the authors employed the original WSI images and the shape before nuclei to obtain two types of coarse labels from the points annotation using the Voronoi diagram and a k-means clustering algorithm. These rough labels are used to train a deep learning model, and then a dense conditional random field is utilized in the loss function to fine-tune the trained model. With a multi-organ

WSI dataset, they achieved a dice score of 0.73. In [12], a conditional generative adversarial network (cGAN) model was proposed for nuclei segmentation, where the segmentation problem was posed as an image-to-image translation task rather than a classification task. A large dataset of synthetic WSI images with perfect nuclei segmentation labels was generated using an unpaired GAN model. Both synthetic and real data with spectral normalization and gradient penalty for nuclei segmentation were used to train the cGAN model.

Zhou et al. [13] presented a deep learning-based model called contour-aware informative aggregation network (CIA-Net) with a multilevel information aggregation module between two task-specific decoders. Instead of using independent decoders, this model exploits bi-directionally aggregated task-specific features to model the spatial and texture dependencies between nuclei and contour. Besides, a smooth truncated loss is utilized to mitigate the perturbation from outliers. As a result, the CIA-Net model is almost built using informative samples, and so its generalization capability could be enhanced (i.e., with multi-organ multi- center nuclei segmentation tasks). With the 2018 MICCAI challenge of the multi-organ nuclei segmentation dataset, they achieved a Jaccard score of 0.63.

Furthermore, the authors of [14] proposed a multi-path dilated residual network for nuclei segmentation and detection. This network comprises the following: 1) a multi-scale feature extraction step based on D-ResNet and feature pyramid network (FPN), 2) a candidate region network, and 3) a final network for detection and segmentation. The segmentation network involves segmentation, regression, and classification sub-networks. With the MonuSeg dataset, they obtained an AJI of 0.46. Mercadier et al. [15] presented a nuclei segmentation framework based on DCNNs. They formulated the problem as segmentation in a holistic manner rather than the classification of patches. The dataset employed is partially annotated, and they used a weighted background model for the network to give more importance to the boundaries of nuclei.

Furthermore, the authors of [16] used a modified version of the U-Net [28] architecture, so-called U-Net++, in which they combined U-Nets of varying depths. With the nuclei segmentation task, they achieved an improvement of 0.0187 with the intersection-over-union (IoU) metric compared to U-Net. The authors of [17] proposed a nuclei segmentation method based on DCNNs for WSI images. To segment nuclei, they used the Mask R-CNN model [18] with color normalization. In particular, the method includes three major steps: preprocessing, nuclei segmentation, and postprocessing. In the preprocessing step, they applied several augmentation techniques to increase the amount of training data and used a color normalization method to reduce the color variation in WSI images. For nuclei segmentation, they followed the implementation of the Mask R-CNN framework stated in the original paper [18] for the backbone network and employed a feature pyramid network (FPN). In the postprocessing, they applied multiple inference methods to improve the segmentation results, obtaining an F1-score of 0.91.

It is worth noting that color variation in histopathological images, the overlapping and clumped nuclei, and the ambiguous boundary between different cell nuclei limit the performance of the above-mentioned nuclei-segmentation methods. Besides, most of the models proposed for this task are complex and do not give the required results. In this study, we present promising deep semantic nuclei segmentation models for overcoming the above-mentioned limitations. To demonstrate the potency of these models, we consider WSI images collected from different scanners of different organs, namely breast, kidney, colon, stomach, prostate, liver, and bladder.

III. METHODOLOGY

A. Nuclei Segmentation Framework

Several deep learning-based semantic segmentation approaches have been proposed in the last decade, supported by the outstanding ability of convolutional neural networks (CNN) in producing semantic and hierarchical image features [19]. In our study, we choose five of the most popular models used for semantic segmentation and adapt them to the nuclei cell segmentation task. Fig. 2 presents the framework of nuclei segmentation, which consists of training and testing phases.

As shown in Fig. 2 (a), the training phase includes a preprocessing step, training the segmentation model, and a postprocessing step. In the preprocessing step, we apply the sparse stain color normalization method of [20] to reduce the variability of color between multi-institutional and multi-organs WSI images. The size of WSI images is high (1000×1000 pixels), and thus it very difficult to train deep learning models with such image size. Thus, we split each WSI into four non-overlapped sub-images. In the postprocessing stage, we assemble the segmented masks corresponding to the four non-overlapped sub-images to restore the original image size.

In the training step, we train different deep learning-based segmentation models. To mitigate overfitting and enhance generalization of the deep learning models, we employ data augmentation techniques. Specifically, we randomly crop 200 patches from training images to augment the data. The number of cropped patches is empirically tuned.
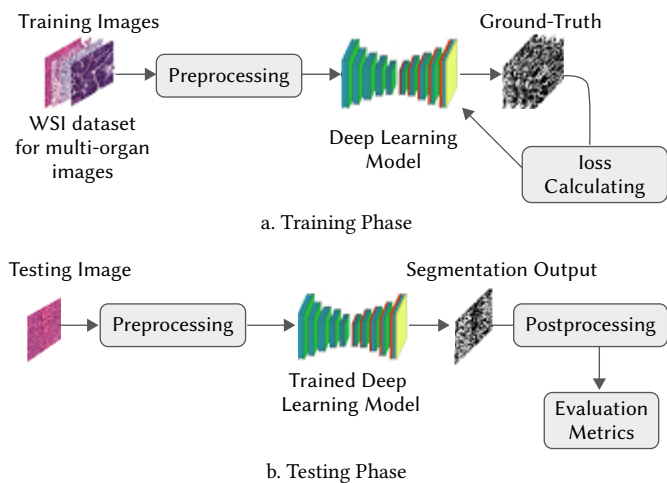

Fig. 2. Nuclei segmentation framework: a) training phase, and b) testing phase.

As shown in Fig. 2 (b), in the test phase we also employ the preprocessing step to normalize the stain of test images and split each test image into four non-overlapped images. Of note, preprocessing step is important to make the same setting used to train the models. After we get the segmented image from a trained model, we apply postprocessing operations to restore the original image size. Besides, we use a connected component algorithm to detect cells and count them. Finally, we evaluate the performance of the nuclei segmentation models in terms of pixel-level F1-score (Dice score) and object-level AJI score metric.

B. Deep Learning-based Semantic Segmentation Models

Fully convolutional network (FCN): In [21], Long et al. proposed the FCN architecture that receives an input image with arbitrary size and produces pixel-wise predictions (i.e. segmentation mask), as shown in Fig. 3. They demonstrated that end-to-end and pixels-to-pixels deep convolutional networks could deliver promising results

with the semantic segmentation task. FCN includes deconvolutional layers to up-sample coarsely deep convolutional layer outputs to dense pixels of any desired resolution.
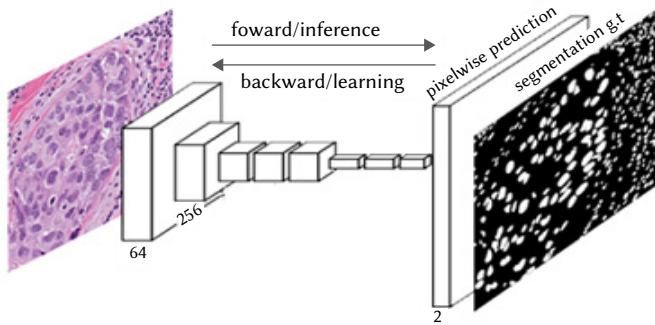

Fig. 3. Diagram of the FCN architecture.

The main idea behind FCN is to create a semantic segmentation network by adjusting state-of-the-art classification networks as such AlexNet [22], the VGG net [23], and GoogLeNet [24] into fully convolutional networks and transfer their representations to the segmentation task (e.g. use of fine-tuning techniques). It is worth noting that the structure of FCN allows generating segmentation maps for images of any resolution without employing fully connected layers, and therefore the FCN architecture is considered as one of the most innovative deep learning architecture that opened the door for several innovations in image segmentation based on deep learning [19]. Besides, skip connections are used to combine semantic information of different layers to produce accurate and detailed segmentation results. Specifically, skip connections enable information to flow, avoiding information loss because of other elements on deep learning architectures, such as max-pooling (down-sampling) and dropout layers. The common FCN architectures are FCN-32, FCN16, and FCN8 [21], which are based on VGG-16 backbone [23]. In our study, we use FCN8 as it gives the best performance.

DenseNet: Several subsequent approaches to semantic segmentation have been inspired by FCN [21], of them the architecture of FC-DenseNets [25]. In short, Jégou et al. [25] extended the architecture

of densely connected convolutional networks (DenseNets) [26], which has achieved remarkable results on image classification tasks, to the semantic segmentation problem. In DenseNet, each layer is connected to other layers in a feed-forward fashion to facilitate the training process. Also, a feature reuse approach is implemented to enable all layers to access their preceding layers.

Fig. 4 (a) shows the architecture of FC-DenseNet, which consists of the down-sampling path described in [26] and the upsampling path which allows recovering the full resolution of input images. As shown, in the down-sampling path the input to a dense block is concatenated with its output, which yields a linear growth in the number of feature maps. Notably, in the down-sampling path, the increase in the number of features is recompensed by decreasing in spatial resolution of each feature map after the pooling operation.

Fig. 4 (b) presents the architecture of the dense block. Of note, in FC-DenseNets an up-sampling process referred to as transition up. Transition up modules consist of a transposed convolution that up-samples the previous feature maps. The up-sampled feature maps are then concatenated to the ones coming by skip connection to form the input of a new dense block.

U-Net: Ronneberger et al. [28] proposed the U-Net architecture, which also is inspired by FCN architecture. The core idea of U-Net and its training strategy is based on the use of data augmentation methods to effectively learn from the available annotated samples. As shown in Fig. 5, the U-Net architecture is built based on the scheme of encoder-decoder networks, which enable capturing the contextual features from input images. The down-sampling path of U-Net (encoder network) follows the typical architecture of FCN to extract features. At each down-sampling step, the number of feature channels is doubled. The up-sampling path (decoder network) consists of deconvolution layers. Feature maps from the encoder network are concatenated with the corresponding ones of the decoder network to avoid losing pattern information (spatial information). Finally, a 1x1 convolution layer is used to generate the segmentation mask, where each pixel of the input channel is assigned to one of the classes.

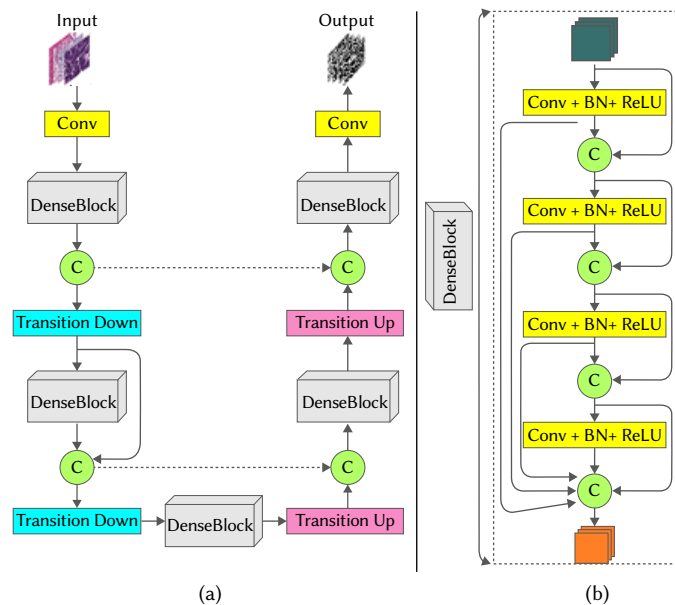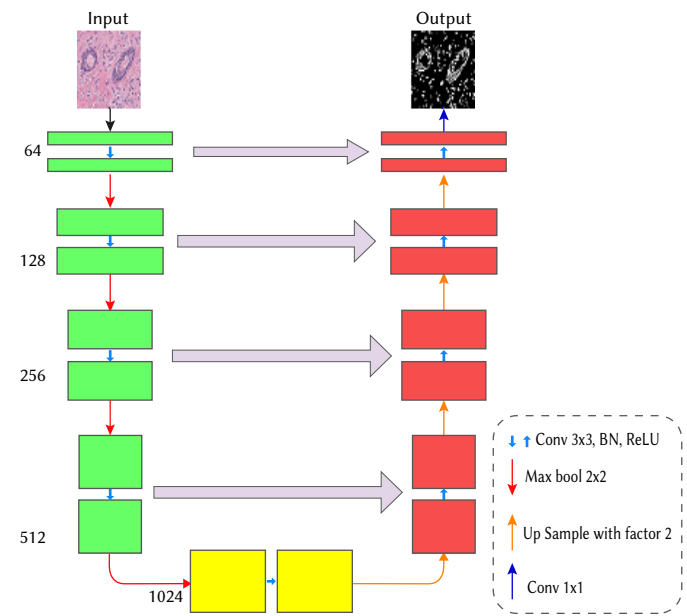Fig. 5. Diagram of the U-Net architecture.

Fig. 4. Diagram of the FC-DenseNets architecture: (a) the two paths of FC-DenseNets architecture for semantic segmentation, and (b) a dense block with 4 layers.

SegNet: In [29], Badrinarayanan et al. presented the SegNet architecture, which consists of an encoder network and a decoder network followed by a pixel-wise classification layer, as shown in Fig. 6. The encoder network of SegNet is similar to the first 13 convolutional

layers of VGG16 [23] without fully connected layers. The decoder network of SegNet comprises a hierarchy of decoders, where each one corresponds to an encoder layer. The decoder layers use max-pooling indices received from the corresponding encoder layers to perform non-linear up-sampling of the feature maps, which eliminates the need for learning to up-sample. The up-sampled maps are then convolved with trainable filters to produce dense feature maps.

Although U-Net and SegNet have similar architecture, U-Net does not reuse pooling indices, and it transfers the entire feature map to the corresponding decoder layers and concatenates them with the up-sampled feature maps, which costs more memory.
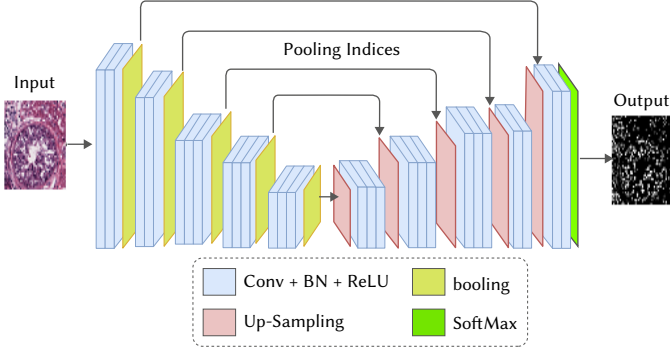

Fig. 6. Diagram of the SegNet architecture.

Self-correction mechanism with CE2P network: In [30], Li et al. introduced a training strategy called self-correction for human parsing (SCHP), which can iteratively improve the reliability of supervised labels as well as the learned models during the training process. The architecture used in SCHP is inspired by the CE2P architecture used in [31]-[32]. It consists of three main branches, namely parsing, edge, and fusion. The training strategy can be divided into two procedures: model aggregation, and label refinement (self-correction mechanism). A cyclically learning scheduler is used to produce reliable pseudo masks. Self-correction is performed iteratively by aggregating the current learned model with the former optimal one in an online manner.

In this study, a cyclically learning scheduler with warm restarts is used. In each cycle of the self-correction mechanism, we compute a set of weights (models), $W = \{\hat{w}_0, \hat{w}_1, \dots, \hat{w}_M\}$ and the corresponding predicted labels, $Y = \{\hat{y}_0, \hat{y}_1, \dots, \hat{y}_M\}$. After each training cycle, the current model weights $\hat{w}$ are combined with the weights of the previous cycle $\hat{w}_{n-1}$ to obtain new weights $\hat{w}_n$, as follows:

$$\hat{w}_n = \frac{n}{n+1}\hat{w}_{n-1} + \frac{1}{n+1}\hat{w} \tag{1}$$

Likewise, the predicted labels of the current cycle are combined with the labels of the previous cycle, as follows:

$$\hat{y}_n = \frac{n}{n+1}\hat{y}_{n-1} + \frac{1}{n+1}\hat{y} \tag{2}$$

where n refers to the current cycle number ($0 \leq m \leq M$) and $\hat{y}$ is the generated pseudo-labels (pseudo masks) with the model $\hat{w}_n$.

PSPSegNet (PSP with SegNet): In [33], Zhao et al. developed the pyramid scene parsing network (PSPNet), which considers the strength of the global context of the image to enhance the local level predictions. The authors of PSPNet claim that FCN based architectures do not employ a suitable strategy to utilize the context of the whole image. Thus, they proposed a pyramid pooling module (PPM) to incorporate global contextual information. For each input image, PSPNet utilizes a pre-trained ResNet (feature extractor) to get feature maps. The feature maps from the feature extractor are pooled at

four different scales corresponding to four different pyramid levels. Then, PPM is used to produce various sub-region representations, succeeded by up-sampling and concatenation layers to produce the final feature maps (i.e. the final representation) that comprise global and local contextual-information. The final representation is inputted into a convolution layer to produce the per-pixel prediction (i.e. the segmentation mask).

To improve the nuclei segmentation results, in this study, we present the PSPSegNet by combining PPM (the key component of PSPNet) with the SegNet architecture, as shown in Fig. 7. We use ResNet [34] as the encoder of SegNet, succeeded by the PPM module. The encoder feature maps are concatenated with the up-sampled outputs of the pyramid levels and then fed into the decoder of the SegNet to produce the segmented image.
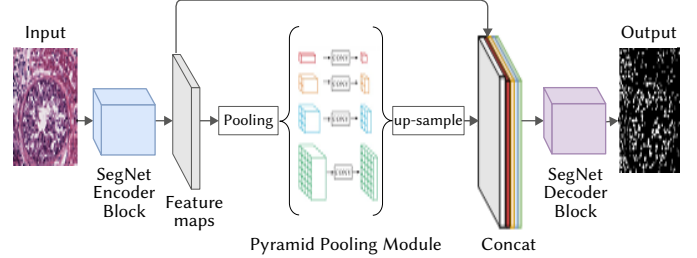

Fig. 7. Diagram of the PSPSegNet architecture.

In this study, we analyze the performance of the five models explained above (FCN, FC-DenseNet, U-Net, Self-correction, and PSPSegNet) with the nuclei segmentation task. We train them using the MoNuSeg dataset [27]. In the test phase, we separately assess the performance of each trained model.

C. Model Evaluation

In this study, we use the aggregated Jaccard index (AJI) proposed in [27] and F1-score (dice score) to assess the performance of the nuclei segmentation methods. AJI is an extended version of the Jaccard index which divides the aggregated intersection cardinality by the aggregated union cardinality between the ground truth (G) and segmented masks. If AJI equals 1, it means that we obtain perfect nuclei segmentation results. AJI can be expressed, as follows:

$$AJI = \frac{\sum_{i=1}^{L}|G_i \cap P_j^*(i)|}{\sum_{i=1}^{K}|G_i \cup P_j^*(i)| + \sum_{K \in Ind}|P_K|} \tag{3}$$

where $G = U_{i=1,2,..K}\ G_i$ is the ground truth masks, $P = U_{i=1,2,..L}\ P_i$ are the prediction nuclei segmentation outputs, $P_j^*(i)$ is the connected component from the prediction output that maximizes the Jaccard index. *Ind* is the list of indices of pixels that do not belong to any element in ground truth (*G*).

The F1-score is the harmonic mean between the precision and recall. The F1-score is identical to the dice coefficient, which can be formulated, as follows:

$$F1 = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{4}$$

where TP, FP, FN refers to the true positive, false positive, and false-negative rates, respectively.

IV. Experimental Results and Discussion

A. Dataset

In our experiments, we use the MoNuSeg dataset [27], which contains 30 WSI images with annotations. MoNuSeg is a very

challenging dataset as it has WSI images of 7 organs (breast, kidney, colon, stomach, prostate, liver, and bladder) collected at different medical centers (i.e., various stains). The size of each WSI image is 1000×1000. Of the 30 WSI images, we use 23 WSI images for training the models and the rest for testing. Mainly, we keep one WSI image for each organ in the testing set.

B. Implementation Details

As introduced in Section III.A, the computation cost of training deep learning models makes it very difficult to train them with the very high resolution of the input image (1000×1000 pixels). Therefore, in the preprocessing step, we apply the sparse stain normalization method of [20] on each WSI image and then rescale it to 1024×1024, and then we divide it into four non-overlapping sub-images of size 512×512. We augment the training data by cropping 200 512×512 patches from each WSI image randomly. Thus, the total training dataset has 4692 of 512×512 patches, and the testing set has 28 sub-images (7 WSI images × 4 splits). A GTX1080 with an 8GB memory GPU is used to run the experiments. All models are trained for 100 epochs, the stochastic gradient descent (SGD) is used as an optimizer

TABLE I. SUMMARIZATION OF MODELS ARCHITECTURE

FCN	FC-DenseNet	U-Net
Input Layer, in_ch= 3		
Feature Extraction Layers 3x3 Conv+ReLU (2 layers) + MaxPool (S= 2), F= 64 3x3 Conv+ReLU (2 layers) + MaxPool (S= 2), F= 128 3x3 Conv+ReLU (3 layers) + MaxPool (S= 2), F= 256 3x3 Conv+ReLU (6 layers) + MaxPool (S= 2), F= 512	**First Layer** 3x3 Conv, F= 48 **Down Sampling Layers** DB (4 layers) + TD, F= 112 DB (5 layers) + TD, F= 192 DB (7 layers) + TD, F= 304 DB (10 layers) + TD, F= 464 DB (12 layers) + TD, F= 656	**Contracting Path Layers** 3x3 Conv+BN+ReLU (2 layers) + MaxPool (S= 2), F=64 3x3 Conv+BN+ReLU (2 layers) + MaxPool (S= 2), F=128 3x3 Conv+BN+ReLU (2 layers) + MaxPool (S= 2), F=256 3x3 Conv+BN+ReLU (2 layers) + MaxPool (S= 2), F=512
Up-sample Layers 3x3 DeConv + BN + ReLU, F= 512 3x3 DeConv + BN + ReLU, F= 256 3x3 DeConv + BN + ReLU, F= 128 3x3 DeConv + BN + ReLU, F= 64 3x3 DeConv + BN + ReLU, F= 32	**Bottleneck Layers** DB (15 layers), F= 896	**Middle Layers** 3x3 Conv+BN+ReLU (2 layers) + MaxPool (S= 2), F=1024
Output (Classifier)_Layer 1x1 Conv, C =2	**Down Sampling Layers** TU + DB (12 layers), F= 1088 TU + DB (10 layers), F= 816 TU + DB (7 layers), F= 578 TU + DB (5 layers), F= 284 TU + DB (4 layers), F= 256	**Expanding Path Layers** Upsample (scale=2) + 3x3 Conv+BN+ReLU (2 layers), F= 512 Upsample (scale=2) + 3x3 Conv+BN+ReLU (2 layers), F= 256 Upsample (scale=2) + 3x3 Conv+BN+ReLU (2 layers), F= 128 Upsample (scale=2) + 3x3 Conv+BN+ReLU (2 layers), F= 64
	Output Layer 1x1 Conv, C = 2 SoftMax	**Output Layer** 1x1 Conv, C = 2

PSPSegNet	SelfCorrection
Input Layer, in_ch= 3	
First Layer 7x7 Conv+BN+ReLU (1 layer) + MaxPool (S= 2), F=64	**First Layers** 3x3 Conv+BN+ReLU (2 layers), F =64 3x3 Conv+BN+ReLU+ MaxPool (S =2), F=128
Encoders Layers (ResNet Backbone) 3 x [(1x1 Conv+BN) + (3x3 Conv+BN)+(1x1 Conv+BN) + ReLU], F= 256 4 x [(1x1 Conv+BN) + (3x3 Conv+BN)+(1x1 Conv+BN) + ReLU], F= 512 23 x [(1x1 Conv+BN) + (3x3 Conv+BN)+(1x1 Conv+BN) + ReLU], F= 1024 3 x [(1x1 Conv+BN) + (3x3 Conv+BN)+(1x1 Conv+BN) + ReLU], F= 2048	**Feature Extraction Layers (ResNet Backbone)** 3 x [(1x1 Conv+BN) + (3x3 Conv+BN)+(1x1 Conv+BN) + ReLU], F= 256 4 x [(1x1 Conv+BN) + (3x3 Conv+BN)+(1x1 Conv+BN) + ReLU], F= 512 23 x [(1x1 Conv+BN) + (3x3 Conv+BN)+(1x1 Conv+BN) + ReLU], F= 1024 3 x [(1x1 Conv+BN) + (3x3 Conv+BN)+(1x1 Conv+BN) + ReLU], F= 2048 **Context_Encoding:** [AvgPool + (1x1 Conv+BN+ReLU)] (4 layers), F=4069 3x3 Conv+BN+ReLU, F= F= 2048
PPM Layer: [AvgPool + (1x1 Conv+BN+ReLU)] (4 layers), F=2048 1x1 Conv, F= 2048	**Parsing Module_Layers** 1x1 Conv+BN+ReLU (4 layers), F =256 1x1 Conv, F= 2
Up-sample Layers 5x5 Conv+BN (4 layers), F=1024, 512, 64, 32 respectively	**Edge_Module Layers** 1x1 Conv+BN (4 layers), F= 256 3x3 Conv, F= 2
Decoder Layers: 1x1 Conv+BN+ReLU, F = 1024 1x1 Conv+BN+ReLU, F = 512 1x1 Conv+BN+ReLU, F = 64 1x1 Conv+BN+ReLU, F = 32	**Output Fusion Module Layers:** 1x1 Conv+BN+ReLU, F=2 1x1 Conv, C= 2
Output_Layers: 1x1 Conv+BN+ReLU, F=2 3x3 Conv, C= 2	

with an initial learning rate of 1e-1, a momentum of 0.99, and a weight decay of 1e-8. A batch size of two images is used.

Table I presents the architecture summarization of FCN, FC-DenseNet, U-Net, PSPSegNet, and Self-Correction models. Where *S*, *F* and *C* stand for stride, the number of feature maps (filters), and the number of output classes, respectively. With the FCN model, we follow the FCN8 architecture presented in [21] and use VGG-16 [23] as a feature extractor with its 13 convolutional layers. We decapitate VGG-16 by discarding the final classifier layer. An up-sampling with 5 deconvolution layers is employed after extracting the features. Finally, we utilize a 1x1 convolution with channel dimension 2 to predict scores for each nuclei class.

In the FC-Dense model, *DB* refers to the Dense-Block shown in Fig. 4. (b), where each layer in the block consists of a 3×3 convolution layer with a batch normalization layer followed by ReLU activation. *TD* refers to transition down operation using a 2×2 max-pooling layer with a stride of 2, and *TU* refers to the transition-up process using a 3x3 transpose convolution layer with a stride of 2.

We follow the same implementation of FC-DenseNet103 architecture presented in [25]. This architecture is built from 103 convolutional layers; the first convolutional layer is applied onto the input, 38 convolutional layers in the down-sampling path, 15 convolutional layers in the bottleneck, and 38 convolutional layers in the up-sampling path. Besides, 5 TD are used, each one containing a convolution, and 5 TU, each one containing a transposed convolution. Finally, a 1×1 convolution layer and a Softmax non-linearity are used to provide the per class distribution at each pixel.

In the U-Net model, two 3×3 convolutional layers followed by ReLU and 2×2 max-pooling operation with stride two are used in each encoder block. In each decoder block, two deconvolution layers are used, which are then concatenated with the corresponding feature maps of the encoder layers. The final layer of the decoder has a 1x1 convolution to map each 64-feature vector to two classes.

In the case of the PSPSegNet model, we use ResNet-101 architecture in the encoder. We implement a PPM between the encoder and the decoder. A bilinear up-sampling operation that consists of four 5×5 convolution layers with batch normalization is applied after PPM. The decoder includes four layers, where each consists of 1×1 convolution and batch normalization layers. Finally, a 1×1 convolution layer followed by a 3×3 convolution layer is used to provide the per class distribution at each pixel.

In the Self-Correction model, we use ResNet-101 [34] as a backbone of the feature extractor and use an ImageNet [35] pre-trained weights to commit with the same implementation in [30]. We adopt the PSP

network [33] as a context encoding module. The parsing module and edge module comprise four 1×1 convolution layers with batch normalization and ReLU followed by one 1×1 convolution layer for the parsing module and one 3×3 convolution layer for the edge module. Finally, a fusion module is employed, which includes a 1×1 convolution layer with batch normalization and ReLU followed by a 1×1 convolution layer to predict scores for each nuclei class. Table II presents the architecture and the source code links of each model.

TABLE II. Models Architecture Backbones

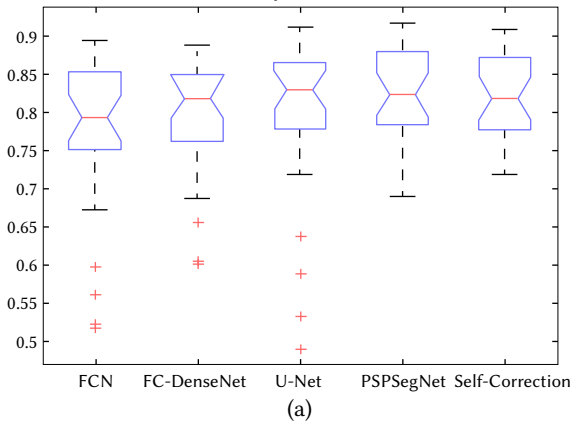| Model	Backbone	GitHub links
FCN	VGG-16	*https://github.com/pochih/FCN-pytorch*
FC-Dense Net	-	*https://github.com/bfortuner/pytorch_tiramisu*
U-Net	-	*https://github.com/LeeJunHyun/Image_Segmentation*
PSPSegNet	ResNet-101	*https://github.com/alexgkendall/SegNet-Tutorial*
Self-Correction	ResNet-101	*https://github.com/PeikeLi/Self-Correction-Human-Parsing*

C. Results and Discussion

Table III shows the segmentation results of the nuclei segmentation models with the MoNuSeg dataset in terms of the dice coefficient (F1-score) and AJI score. As shown, FCN obtains F1-score of 0.8467 and AJI of 0.6418, which are lower than the other four models. FC-Dense Net and U-Net achieve improvements on the F1-score of 1.2% and 1.56%, respectively, when compared to FCN. Besides, they give gains on the AJI of 2.28% and 3.38%, respectively. PSPSegNet obtains the best results with 3.48% improvement on the F1-score and 6.62% improvement on AJI, thanks to PPM that encourages the PSPSegNet model to learn global context features of WSI images. The F1-score and AJI of PSPSegNet are 0.26% and 0.66% higher than the ones of Self-Correction. This analysis reveals that both PSPSegNet and Self-Correction could be used to get suitable nuclei segmentation results.

TABLE III. Comparison Between the Nuclei Segmentation Models

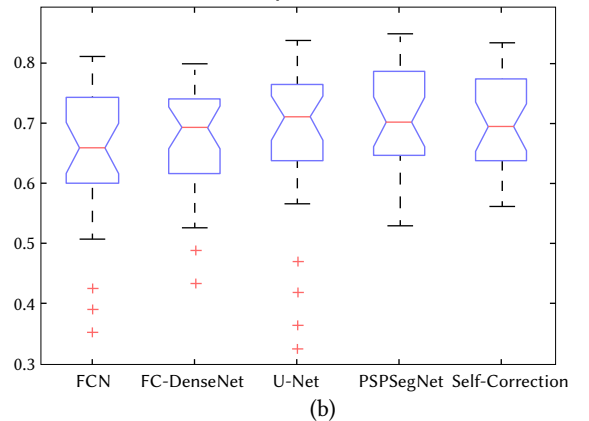|	F1-Score	AJI
FCN	0.8467	0.6418
FC-Dense Net	0.8587	0.6646
U-Net	0.8623	0.6756
PSPSegNet	**0.8815**	**0.7080**
Self-Correction	0.8792	0.7014

Fig. 8. Boxplots of F1-score and AJI of the five nuclei segmentation models: (a) F1-score, and (b) AJI.

Fig. 8 shows the boxplots of F1-score and AJI for all nuclei segmentation models. Given the scores of test images with a particular model, a boxplot can be displayed based on a five-number summary: the minimum, the maximum, the sample median, and the first and third quartiles. In Fig. 8, the red horizontal line refers to the sample median. As shown, the PSPSegNet and Self-Correction models have not any outliers on F1-score and AJI values. U-Net model has the highest median F1-score and AJI, but it produces the highest number of outliers on both evaluation metrics. However, FCN has AJI less than U-Net; it has a lower number of outliers. As we can see, PSPSegNet and Self-Correction models almost achieve the same median AJI and F1-score values. PSPSegNet achieves the maximum AJI and F1-scores when compared to other models, while FCN produces the minimum values.

Fig. 9 presents samples of segmented WSI images of different organs. We can notice that segmentation results can vary from one organ to another. For example, the WSI image of the liver organ (Fig. 9 (a)) has several big nuclei and some of them are overlapped. As shown in Fig. 9 (a) Col. 6, the PSPSegNet model accurately segments the cell nuclei with an AJI score 0.635, while the FCN model gives the worst segmentation results with an AJI score of 0.349 (Fig. 9 (a), Col. 3). We believe that the good performance of PSPSegNet is a result of the employment of PPM that integrates multi-scale maps in the middle of the model to learn WSI image context features. The same conclusion can be said for the colon organ WSI image (Fig. 9 (c)) and the bladder WSI image (Fig. 9 (e)).

(a) Liver.

(b) Breast.

(c) Colon.

(d) Kidney.
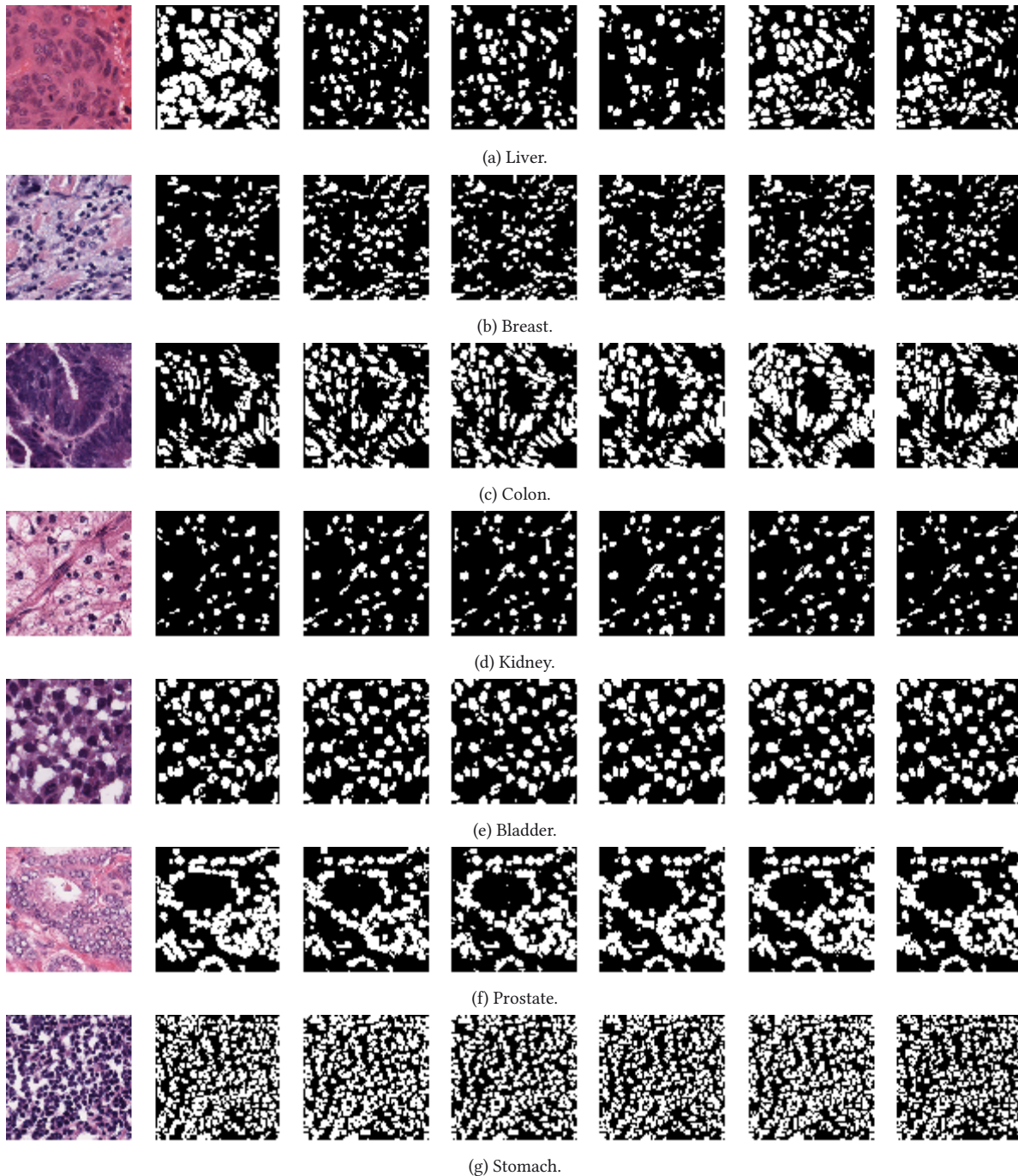
(e) Bladder.

(f) Prostate.

(g) Stomach.

Fig. 9. Segmentation results of all models with different organs (liver, breast, colon, kidney, prostate, bladder, and stomach). The first and second columns represent the original input image and ground truth mask, respectively. Columns 3-7 represent the output mask of FCN, FC-DenseNet, U-Net, PSPSegNet, and SelfCorrection, respectively.

Fig. 9 (b) shows a WSI image of the breast organ that has a noticeable color stain variation with a lot of overlapped nuclei. As shown in Fig. 9 (b) Col. 7, the Self-Correction model obtains the best performance with an AJI score of 0.655, thanks to the cyclically learning scheduler that enhances the segmentation results. The same conclusion can be said for the WSI image of the kidney organ shown in Fig. 9 (d). Fig. 9 (g) presents a WSI image of the stomach organ that has a dense number of nuclei. As we can see, all models produce good segmentation results. Specifically, the PSPSegNet model achieves the best results with an AJI score of 0.829. It is worth noting that all models obtain good segmentation results with stomach, liver and breast WSI images. However, in the case of the most complex WSI images that have color stain variation and overlapped nuclei, PSPSegNet and Self-Correction models produce the best segmentation results.

As nuclei segmentation is crucial to cell counting, it is interesting to study the performance of the five segmentation models with this task. In this regard, Fig. 10 presents a comparison between the number of cell nuclei in the predicted masks of each segmentation model and the ground-truth. To count the number of cell nuclei in each mask, we employ the connected component algorithm. In this experiment, we empirically set a threshold of 70 pixels for the minimum area of cells. As shown in Fig. 10, the number of cell nuclei obtained by the PSPSegNet model is a bit higher than the ones of the ground-truth, while the number of cell nuclei obtained by the Self-Correction model is close to the ground-truth. In turn, FCN, FC-Dense, and U-Net models have a lower number of cells than the ground-truth.

Fig. 11 shows a comparison between the number of parameters of FCN, FC-DenseNet, U-Net, PSPSegNet, and self-correction models. As shown, there is a noticeable variation in complexity between the nuclei segmentation models. It is worth noting that the good results achieved with the PSPSegNet model (F1-score 0.882 and AJI score 0.708) cost a massive number of parameters that exceed 122 million.
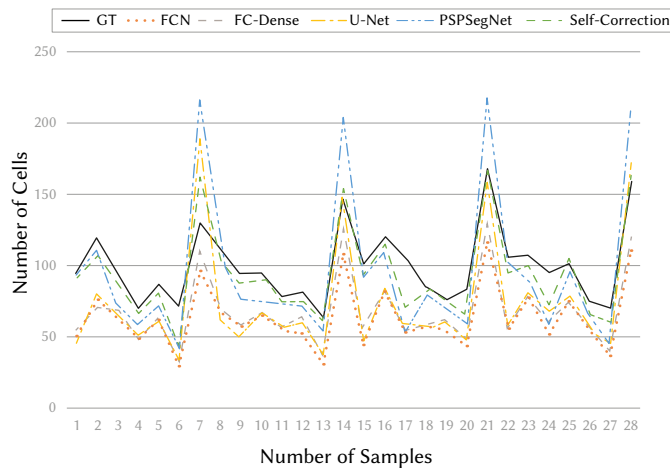

Fig. 10. Comparison between the number of cell nuclei in the predicted masks of each segmentation model and the ground-truth.

The Self-Correction model has 66 million parameters, which is almost half of PSPSegNet while achieving acceptable segmentation results (F1-score 0.879 and AJI score 0.701). However, the number of parameters of the FC-DenseNet model is less than the FCN model (around 9 million); it achieves a better F1-score and AJI. The FC-DenseNet model is also better than the U-Net model in terms of the number of parameters without a big difference in the segmentation results.
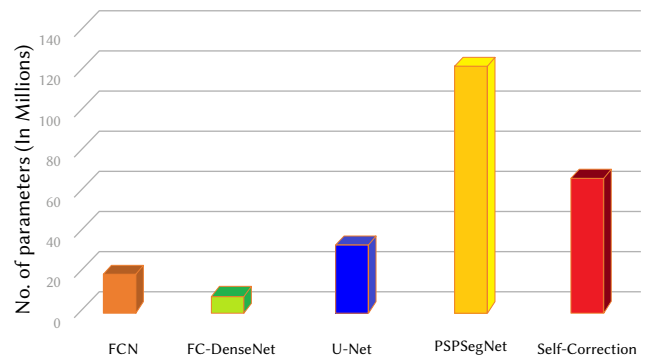

Fig. 11. Comparison between the number of parameters of FCN, FC-DenseNet, U-Net, PSPSegNet, and self-correction models.

In this study, we also ensemble the prediction masks of the top three models (U-Net, PSPSegNet, and self-Correction) using a simple pixel-wise aggregation function, as follows:

$$Ens = OR\,[U_{mask}, AND\,(PSP_{mask}, S_{mask})] \tag{5}$$

where U_{mask}, PSP_{mask}, and S_{mask} are the prediction masks of U-Net, PSPSegNet, and Self-Correction models, respectively. *Ens* is the ensemble output. This ensemble method increases the AJI score to 0.7103 (0.23% improvement on the AJI compared to PSPSegNet).

Indeed, pathologists prefer to use easy and friendly software to segment cells from histopathology images. One of the most popular software is ImageFIJI [3]. Fig. 12 shows the segmentation results of the ImageFIJI program. As shown, the segmentation result of ImageFIJI is worse than the one of the proposed ensemble model (U-Net, PSPSegNet, and Self-Correction models). Also, ImageFIJI gives an AJI score of 0.533, which is much lower than the one of FCN model (the worst nuclei segmentation model presented in this study). Therefore, PSPSegNet could be a proposing nuclei segmentation method for pathologists.
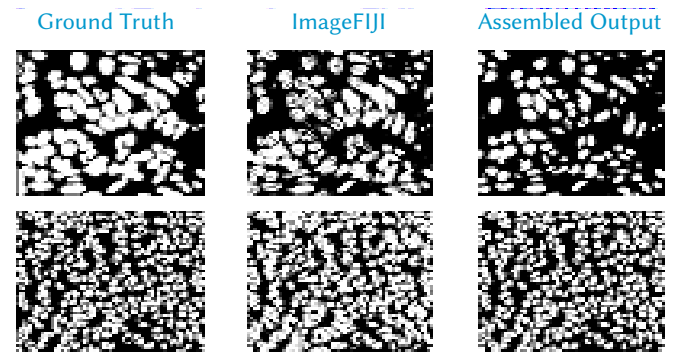

Fig. 12. Segmentation results of (middle) ImageFIJI software, and (right) ensemble of U-Net, PSPSegNet, and Self-Correction models.

In the literature, several studies have also used the MoNuSeg dataset and evaluated the segmentation models using F1-score and AJI. For instance, the study of [12] proposed a deep conditional generative adversarial network (cGAN) model, the study of [11] proposed a weakly supervised deep nuclei segmentation model, and the study of [7] presented the bending loss regularized network, to segment nuclei in WSI images. Mahmood et al. [12] obtained an F1 score of 0.866 and an AJI score 0.721, Qu et al. [11] achieved an F1 score of 0.778 and an AJI score of 0.505, and Wang et al. [7] obtained an AJI score 0.641. Based on the results mentioned above, we can conclude that PSPSegNet and self-correction approach outperforms all these approaches in term of pixel-level F1-score. Besides, the results achieved in [11] are much lower than the ones of PSPSegNet, and Self-

Correction models in terms of F1-score and AJI, noting that weakly supervised nuclei segmentation using points annotation technique may not be appropriate for cell nuclei segmentation. In terms of object-level AJI score, we can see that the results of [12] exceed our models, noting that this approach depends on synthetic training data that may not produce very accurate cell shapes.

To end, the models presented in this study can overcome the challenges that existing nuclei methods face. Specifically, PSPSegNet achieves promising performance in terms of the F1-score AJI scores. Thus, PSPSegNet could be a feasible nuclei segmentation tool for pathologists. It is worth noting that the segmentation model presented in this paper can be used to segment region of interest in several medical image modalities, such as the segmentation of nipples in thermograms [35], segmentation of pectoral muscle in mammograms [36], and vessel segmentation in fundus images [37].

V. Conclusion

In this paper, we have presented promising deep semantic nuclei segmentation models in WSI images of different organs and collected from various clinics. To overcome the challenges that existing nuclei segmentation models face, we have sought the efficacy of pertinent deep learning models with nuclei segmentation task. Besides, we have consolidated robust deep learning architectures to build an efficient deep learning nuclei segmentation model (named PSPSegNet).

To demonstrate the performance of the nuclei segmentation models, we have used a well-known multi-organ WSI image dataset that includes WSI collected from different organs and scanners. We have comprehensively compared the performance of the nuclei segmentation models and exiting software in terms of the F1-score metric, object-level AJI metric, and the number of trainable parameters. The proposed PSPSegNet model achieved the highest performance with a pixel-level F1-score of 0.8815 and an object-level AJI score of 0.7080. PSPSegNet achieves promising results, but it has 122 million trainable parameters. In comparison with FCN, PSPSegNet achieved 3.48% improvement on the F1-score and 6.62% improvement on AJI.

Interesting results have been obtained with the Self-correction model with an F1-score of 0.8815 and AJI-score of 0.7080 with almost 67 million trainable parameters. Of note, the number of trained parameters of FCN, FC-DenseNet, and U-Net models ranges from 9 to 34 million, but they obtained lower segmentation performance than PSPSegNet with F1-scores of 0.847, 0.859, 0.862 respectively, and AJI scores of 0.641, 0.665, 0.676 respectively. Also, we have compared the performance of PSPSegNet with existing software (ImageFIJI), noting that PSPSegNet achieves better results. The experimental results emphasize that the PSPSegNet model could be used in the cell counting task.

The future work will include several extensions of the current study:

- The use of different aggregation strategies to combine the individual nuclei segmentation models.
- Incorporation of stain normalization techniques into the deep learning framework with different strategies.
- Converting the WSI images to other coordinate systems, such as the log-polar coordinates [39] and the curvilinear coordinates [37] to improve the nuclei segmentation results.

Acknowledgment

References

[1] Moen, E., Bannon, D., Kudo, T. et al. "Deep learning for cellular image analysis," i*n Nat Methods*, vol. 16, no. 12, pages. 1233–1246, 2019, doi: 10.1038/s41592-019-0403-1.

[2] Carpenter, A.E., Jones, T.R., Lamprecht, M.R. et al. "CellProfiler: image analysis software for identifying and quantifying cell phenotypes," *in Genome Biol*, vol 7, R100, 2006, doi: 10.1186/gb-2006-7-10-r100.

[3] Schindelin, J., Arganda-Carreras, I., Frise, E. et al., "Fiji: an open-source platform for biological-image analysis," *in Nat Methods*, vol. 9, no. 7, pages. 676- 682, 2012.

[4] Niazi, Muhammad Khalid Khan et al. "Digital pathology and artificial intelligence," *in The Lancet. Oncology*, vol. 20, no. 5, pages e253-e261, 2019, doi: 10.1016/S1470-2045(19)30154-8

[5] P. Naylor, M. Laé, F. Reyal and T. Walter, "Nuclei segmentation in histopathology images using deep neural networks," *in 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, Melbourne, VIC, 2017, pp. 933-936, doi: 10.1109/ISBI.2017.7950669.

[6] P. Naylor, M. Laé, F. Reyal and T. Walter, "Segmentation of Nuclei in Histopathology Images by Deep Regression of the Distance Map," *in 2019 IEEE Transactions on Medical Imaging*, vol. 38, no. 2, pp. 448-459, Feb. 2019, doi: 10.1109/TMI.2018.2865709.

[7] H. Wang, M. Xian and A. Vakanski, "Bending Loss Regularized Network for Nuclei Segmentation in Histopathology Images," *in 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, Iowa City, IA, USA, 2020, pp. 1-5, doi: 10.1109/ISBI45749.2020.9098611.

[8] Graham, Simon Matthew, Quoc Dang Vu, Shan-e-Ahmed Raza, Jin Tae Kwak and Nasir M. Rajpoot. "XY Network for Nuclear Segmentation in Multi-Tissue Histology Images," in ArXiv abs/1812.06499 2018.

[9] Al-Kofahi, Y., Zaltsman, A., Graves, R. et al. "A deep learning-based algorithm for 2-D cell segmentation in microscopy images," *in BMC Bioinformatics*, vol. 19, no. 365, 2018, doi: 10.1186/s12859-018-2375-z

[10] Cui, Y., Zhang, G., Liu, Z. et al. "A deep learning algorithm for one-step contour aware nuclei segmentation of histopathology images," *in Med Biol Eng Comput*, vol. 57, pages 2027–2043, 2019, doi: 10.1007/s11517-019-02008-8

[11] H. Qu et al., "Weakly Supervised Deep Nuclei Segmentation Using Partial Points Annotation in Histopathology Images," *in 2020 IEEE Transactions on Medical Imaging*, doi: 10.1109/TMI.2020.3002244.

[12] F. Mahmood et al., "Deep Adversarial Training for Multi-Organ Nuclei Segmentation in Histopathology Images," *in 2019 IEEE Transactions on Medical Imaging*, 2019, doi: 10.1109/TMI.2019.2927182.

[13] Zhou, Y., Onder, O.F., Dou, Q., Tsougenis, E., Chen, H., Heng, P.A., "CIA-Net: Robust Nuclei Instance Segmentation with Contour-Aware Information Aggregation," *International Conference on Information Processing in Medical Imaging*, 2019, pp. 682-693.

[14] Wang, E.K., Zhang, X., Pan, L., Cheng, C., Dimitrakopoulou-Strauss, A., Li, Y., Zhe, N., "Multi-Path Dilated Residual Network for Nuclei Segmentation and Detection," *Cells Journal, Multidisciplinary Digital Publishing Institute*, 2019, vol. 8, no. 5, page 499.

[15] D. S. Mercadier, B. Besbinar and P. Frossard, "Automatic Segmentation of Nuclei in Histopathology Images Using Encoding-decoding Convolutional Neural Networks," *in 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, United Kingdom, 2019, pp. 1020-1024, doi: 10.1109/ICASSP.2019.8682502.

[16] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh and J. Liang, "UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation," in *2020 IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1856-1867, June 2020, doi: 10.1109/TMI.2019.2959609.

[17] Jung, H., Lodhi, B., Kang, J., "An automatic nuclei segmentation method based on deep convolutional neural networks for histopathology images," *BMC Biomedical Engineering Journal*, vol.1, no. 24, 2019, doi: 10.1186/s42490-019-0026-8.

[18] K. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask R-CNN," *in 2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, 2017, pp. 2980-2988, doi: 10.1109/ICCV.2017.322.

[19] Lateef, Fahad, and Yassine Ruichek. "Survey on semantic segmentation using deep learning techniques," *in Neurocomputing*, vol. 338, pp. 321-348, 2019.

[20] A. Vahadane, T. Peng, A. Sethi, S. Albarqouni, L. Wang, M. Baust, K.

Steiger, A. M. Schlitter, I. Esposito, and N. Navab, "Structurepreserving color normalization and sparse stain separation for histological images," *in 2016 IEEE transactions on medical imaging*, vol. 35, no. 8, pp. 1962–1971, 2016.

[21] J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation," *in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, 2015, pp. 3431-3440, doi: 10.1109/CVPR.2015.7298965.

[22] Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. "ImageNet Classification with Deep Convolutional Neural Networks," *Neural Information Processing Systems*, vol. 25, 2012.

[23] Simonyan, Karen and Zisserman, Andrew. "Very Deep Convolutional Networks for Large-Scale Image Recognition," *in Computing Research Repository (CoRR)*. [Online]. Available: http://arxiv.org/abs/1409.1556.

[24] C. Szegedy et al., "Going deeper with convolutions," *in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, 2015, pp. 1-9, doi: 10.1109/CVPR.2015.7298594.

[25] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero and Y. Bengio, "The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation," *in 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, HI, 2017, pp. 1175-1183, doi: 10.1109/CVPRW.2017.156.

[26] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, "Densely Connected Convolutional Networks," *in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 2017, pp. 2261-2269, doi: 10.1109/CVPR.2017.243.

[27] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane and A. Sethi, "A Dataset and a Technique for Generalized Nuclear Segmentation for Computational Pathology," in 2017 *IEEE Transactions on Medical Imaging*, vol. 36, no. 7, pp. 1550-1560, July 2017, doi: 0.1109/TMI.2017.2677499.

[28] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation", *International Conference on Medical image computing and computer-assisted intervention*. Springer, vol 9351, pp. 234–241, 2015, doi: 0.1007/978-3-319-24574-4_28.

[29] V. Badrinarayanan, A. Kendall and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *in 2017 IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481-2495, 1 Dec. 2017, doi: 10.1109/TPAMI.2016.2644615.

[30] H. Zhao, J. Shi, X. Qi, X. Wang and J. Jia, "Pyramid Scene Parsing Network," *in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 2017, pp. 6230-6239, doi: 10.1109/CVPR.2017.660.

[31] Li, P., Xu, Y., Wei, Y., and Yang, Y. "Self-correction for human parsing," 2019, arXiv:1910.09777.

[32] Ruan, T., Liu, T., Huang, Z., Wei, Y., Wei, S., and Zhao, Y., "Devil in the Details: Towards Accurate Single and Multiple Human Parsing". *In 2019 Proceedings of the AAAI Conference on Artificial Intelligence*, 2019 volume 33, pages 4814–4821.

[33] Abdel-Nasser, M., Saleh, A., and Puig, D. "Channel-wise Aggregation with Self-correction Mechanism for Multi-center Multi-Organ Nuclei Segmentation in Whole Slide Imaging". *In 2020 Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, VISIGRAPP 2020, Vol. 4: VISAPP, Valletta, Malta, February 27-29, 2020 (pp. 466–473).

[34] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[35] J. Deng, W. Dong, R. Socher, L. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," *in 2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, 2009, pp. 248-255.

[36] Abdel-Nasser, M., Saleh, A., Moreno, A., and Puig, D. "Automatic nipple detection in breast thermograms," *Expert Systems with Applications*, vol. 64, pp. 365-374, 2016, doi: 10.1016/j.eswa.2016.08.026

[37] Abdel-Nasser, M., Moreno, A., and Puig, D. "Temporal mammogram image registration using optimized curvilinear coordinates," *Computer methods and programs in biomedicine*, vol. 127, pp. 1-14, 2016. doi: 10.1016/j.cmpb.2016.01.019

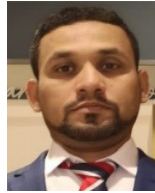[38] WANG, Xiaohong; JIANG, Xudong; REN, Jianfeng. "Blood vessel segmentation from fundus image by a cascade classification framework," *in Pattern Recognition*, 2019, vol. 88, p. 331-341.

[39] Roberts, N., Magee, D., Song, Y., Brabazon, K., Shires, M., et. al. "Toward routine use of 3D histopathology as a research tool," *The American journal of pathology*, vol. 180, no. 5, pp. 1835-1842, 2012.

Loay Hassan

He received the B.Sc. and M.Sc. degrees from Aswan University, in 2010 and 2015, respectively. Currently, he is a PhD student in Aswan university. He worked as an assistant researcher at Center for Artificial Intelligence and Robotics (CAIRO), Aswan University for 4 years. Currently, he is working as a part-time teacher assistant in Department of computer science at Arab Academy for Science, Technology & Maritime Transport, south valley branch.

Adel Saleh

He is currently a researcher at Gaist Solutions Ltd (UK). He received his Ph.D. in Computer Engineering from Universitat Rovira i Virgili (Spain) in 2019 and his master degree in computer science from VSTU (Russia). Dr. Saleh has published more than 20 papers in international journals and conferences. His research interests include images and videos analysis based on deep learning.

Mohamed Abdel-Nasser

He received the Ph.D. degree in Computer Engineering from the University Rovira i Virgili (URV) in 2016. He has been a postdoc researcher at the URV since 2018, and Assistant Professor at the Electrical Engineering Department, Aswan University (Egypt) since 2016. In 2017, he has received the Marc Esteva Vivanco prize for the best Ph.D. dissertation on Artificial Intelligence. He has participated in several projects funded by the European Union and the Government of Spain. He has published more than 65 papers in international journals and conferences. He was the manager of the E-learning & digital library center at Aswan University (2018). His research interests include the application of machine learning and deep learning to several real-world problems, including medical image analysis, smart road environment, smart grid analysis and time-series forecasting.

Osama A. Omer

He received the B.Sc. and M.Sc. degrees from South Valley University, in 2000 and 2004, respectively, and the Ph.D. degree from the Tokyo University of Agriculture and Technology, in 2009. He spent six months as a Postdoctoral Researcher with the Medical Engineering Department, Luebeck University, Germany. Also, he spent three months as a Postdoctoral Researcher with Kyushu University, Japan. He spent six months as an R&D Scientist Engineer with the NOKIA R&D Center, Tokyo, Japan, in 2008. He is currently a Professor with the Electrical Engineering Department, Aswan University. Prof Omer has published more than 70 papers in international journals and conferences. His research interests include medical imaging, super-resolution, image/video coding, and wireless communications.

Domenec Puig

He received the M.S. and Ph.D. degrees in computer science from the Polytechnic University of Catalonia, Barcelona, Spain, in 1992 and 2004, respectively. In 1992, he joined the Department of Computer Science and Mathematics, Rovira i Virgili University (URV), Tarragona, Spain, where he is currently working as a Professor since 2017. He has been the Head of the Intelligent Robotics and Computer Vision Group, Rovira i Virgili University since 2006. Prof. Puig is also the Vice-Rector of the URV. He is the principal investigator (PI) of several projects funded by the European Union and the Government of Spain. He has published more than 200 papers in international journals and conferences His research interests include artificial intelligence and mobile robotics.