



# High-resolution deep transferred ASPPU-Net for nuclei segmentation of histopathology images

Amit Kumar Chanchal<sup>1</sup> · Shyam Lal<sup>1</sup> · Jyoti Kini<sup>2</sup>

Received: 3 February 2021 / Accepted: 8 September 2021 / Published online: 7 October 2021  
© CARS 2021

## Abstract

**Purpose** Increasing cancer disease incidence worldwide has become a major public health issue. Manual histopathological analysis is a common diagnostic method for cancer detection. Due to the complex structure and wide variability in the texture of histopathology images, it has been challenging for pathologists to diagnose manually those images. Automatic segmentation of histopathology images to diagnose cancer disease is a continuous exploration field in recent times. Segmentation and analysis for diagnosis of histopathology images by using an efficient deep learning algorithm are the purpose of the proposed method.

**Method** To improve the segmentation performance, we proposed a deep learning framework that consists of a high-resolution encoder path, an atrous spatial pyramid pooling bottleneck module, and a powerful decoder. Compared to the benchmark segmentation models having a deep and thin path, our network is wide and deep that effectively leverages the strength of residual learning as well as encoder–decoder architecture.

**Results** We performed careful experimentation and analysis on three publically available datasets namely kidney dataset, Triple Negative Breast Cancer (TNBC) dataset, and MoNuSeg histopathology image dataset. We have used the two most preferred performance metrics called F1 score and aggregated Jaccard index (AJI) to evaluate the performance of the proposed model. The measured values of F1 score and AJI score are (0.9684, 0.9394), (0.8419, 0.7282), and (0.8344, 0.7169) on the kidney dataset, TNBC histopathology dataset, and MoNuSeg dataset, respectively.

**Conclusion:** Our proposed method yields better results as compared to benchmark segmentation methods on three histopathology datasets. Visual segmentation results justify the high value of the F1 score and AJI scores which indicated that it is a very good prediction by our proposed model.

**Keywords** Deep learning · Dilated convolution · Histopathology image · Nuclei segmentation

## Introduction

Histopathology is one of the best methods to diagnose cancer disease and early detection of cancer disease is the demand of today's time. A histopathology image consists of hematoxylin stains cell cores called nuclei, and eosin stain cells

with other connected tissue called cytoplasm which is of pink shading. According to the pathologist report [1], grade of cancer is determined by the morphology of tissue while the stage of cancer in the body is decided by tumor size, location, and spread. Most of the information like the stage of cancer, the grade of cancer, mitotic rate, lymph node status of the cancer tissue is identified by analyzing the hematoxylin morphology, location, and spread. The traditional method to diagnose cancer disease is to examine the morphological structure and distribution of nuclei manually by pathologists. After taking the image of the sample tissue, pathologists decide whether tissue regions are cancerous or not and how much is the malignancy level. A lot of time and effort are required in the manual analysis of histopathology images due to the complexity involved in such images. Preparation of histopathology slides is the first

✉ Shyam Lal  
shyam.mtec@gmail.com

Amit Kumar Chanchal  
amit.chanchal01@gmail.com

<sup>1</sup> Department of E&C Engineering, National Institute of Technology Karnataka, Surathkal, Mangaluru, Karnataka 575025, India

<sup>2</sup> Department of Pathology, Kasturba Medical College, Manipal Academy of Higher Education, Mangalore, Manipal, India

stage of the automatic segmentation method which involves tissue collection, fixation, embedding, sectioning, staining, and visualization. The collection of samples from the human body and preserved these samples with some fixation material is the primary step in the preparation of slides. The tissue is implanted in several blocks called embedding and this tissue is used as a block form for the sectioning purpose. Usually, we require a thinner section for the diagnosis purpose. Adding color in the thinner section of tissue is basically to make nuclei more expressive called staining and finally, the stained slide is examined by the different WSI scanners. Most pathologists rely on the visualization of histopathology slides with the help of a microscope or the WSI scanner because a microscope provides faster focusing and scanning. As a segmentation task, our focus is to separate hematoxylin stains cells from eosin stain cells and other connected tissue. Nuclei segmentation can be categorized into conventional and CNN-based deep learning segmentation approaches. The traditional method of image segmentation includes the discontinuity-based approach, similarity-based approach, global clustering, superpixel segmentation, watershed segmentation, active contour segmentation techniques, etc. A summary of conventional image segmentation is presented in Table 1.

These methods performed better in the field of biomedical application but the main challenge is to deal with complex histology images where conventional techniques result in under-segmentation or over-segmentation. To perform the segmentation task, we collected three datasets namely the kidney cancer histopathology image dataset, Triple-Negative Breast Cancer (TNBC) dataset, and Multi-Organ nuclei segmentation (MoNuSeg) dataset. In this study, we are using a deep learning framework for the segmentation process where features are learned automatically from the model, and based on that they predict the segmented image. Deep learning architectures learn the model parameters to separate the nuclei in histology images. The deep learning framework is driven by large amounts of data, activation function, an effective optimization algorithm, and powerful loss function. An activation function can map the input to a meaningful number and an optimization function is responsible for the fast convergence of model parameters. A loss function effectively calculates the difference between the true value and predicted value in the case of regression problems while calculates true probability map and predicted probability map in the case of classification problems. Many recent studies reported that deep learning-based segmentation models provide better accuracy compared to traditional manual feature extraction methods. Although the potential of a deep learning algorithm has proven to be better, these are the challenges in the automatic segmentation of H & E stained histopathology images.

1. Difficult to deal with the clumped morphological appearance of nuclei in histopathology images, as a single histopathological slide may have more than thousands of nuclei.
2. Requires a sufficient amount of high-quality labeled data verified by an experienced medical practitioner.
3. Computational cost is an important issue in deep learning models.

To address the challenges associated with the nuclei segmentation from histopathological images, our contribution is as follows:

1. A most accurate deep learning architecture of nuclei segmentation from H & E stained histopathology images.
2. To enrich the features at five distinct stages, our proposed model effectively utilized residual connection throughout the network and atrous spatial pyramid pooling (ASPP) layer at the bottleneck.
3. Compared to the benchmark segmentation models composed of the deep and thin path, the proposed network is wide and deep that effectively leverages the strength of residual learning as well as encoder–decoder architecture.
4. The proposed model carefully experimented on three publicly available histopathology dataset and outperform five state-of-the-art deep learning models in terms of quality metrics.

The rest of the manuscript is organized as follows: In the related work section, we provide a concise study of the most recent benchmark approach. The detailed mathematical analysis of the proposed model is presented in the proposed architecture section. The process of experimentation is described in the training and implementation section. Experimental results and comparisons are presented in the results and discussion section. Conclusion and future works we provide in the conclusion section.

## Related work

Deep learning techniques are a good option whenever availability of sufficient data. As a segmentation task, to regain the relevant information which we lose while going deeper in the network during pooling operation is the most important. A great contribution in the field of biomedical image segmentation [11] called UNet. UNet is an encoder and decoder path that care about both the context and location of an object in the image. UNet achieves very good prediction result for different segmentation applications. For gland segmentation, a deep contour aware network (DCAN) proposed by Hao Chen et al. [12], able to separate overlapped objects. DCAN

**Table 1** Conventional (Non deep learning) image segmentation methods

Approach	Description	Advantages	Limitations
Discontinuity based approach [2] (Boundary approach)	This method is based on the principle of intensity variations among the pixels	Works well for images having good contrast between regions	Boundaries determined by this method is generally discontinuous
Point detection	Extract regions that differ in properties like intensity, color, texture etc	Second-order derivative gives a reliable result	The single operator does not suit for all types of images
Line detection			
Edge detection			
Similarity based approach [2] (Region approach)	Group those pixels which are similar in some sense	Gives better result compared to other conventional segmentation methods	Expensive in terms of computation time and memory
Region growing	Common properties of the pixels are utilized	Flow from inner point to outer region generates clear object boundaries	Selection of noisy seed leads to wrong segmentation result
Region splitting and merging			
Thresholding method			
Global clustering based approach	In clustering, all pixels in the plane have equal weights	Relatively simple to implement	Sensitive to initial values
<i>K</i> -means [3] clustering	<i>K</i> -means performs grouping based on their similarity by comparing the data with each other	Time taken to cluster <i>K</i> -means rises linearly with the number of data points	Falling easily into local minima
Fuzzy-C means [4] clustering	In fuzzy C-means samples to be clustered belongs to a cluster center in a certain plane		Sensitive to noise
Superpixel (A cluster of connected pixels with similar features like color, brightness, texture, etc.) segmentation methods [5,6]	SLIC superpixel segmentation is used as pre-segmentation to the clustering-based segmentation method	Applicable in MRI image segmentation, Glaucoma screening, Optic disc segmentation	Superpixel methods extract the meaningful regions in the image and improve the computation based on the pixel
Simple linear iterative clustering (SLIC)	SLIC, clusters neighbor pixels by considering their color and coordinate information	Superpixel methods used in histopathological analysis are about the segmentation of regions rather than cell segmentation	The superpixel method with large parameters leads to slow running time and poor segmentation performance
Density-based spatial clustering of application with noise (DBSCAN)	DBSCAN is a clustering algorithm based on the measurement of density in regions close to the given object	SLIC method controls the tradeoff between superpixel compactness and boundary adherence	Different superpixel methods have different advantages and drawbacks. We should choose the proper method according to the problem
Topological preserved regular superpixel (TPRS)	DBSCAN improves the performance of segmentation by adding local geometric boundaries	Provides regional information	
Entropy rate superpixel (ERS)	In TPRS arranged seeds relocated to the pixel with locally maximal edge magnitude	High computational efficiency	

**Table 1** continued

Approach	Description	Advantages	Limitations
Mean Shift	ERS finds compact and homogeneous superpixels by using a graph-based approach	Good adherence to object boundaries	
Quick Shift			
Normalized Cut (Neut)			
Watershed segmentation methods [7,8]	MC watershed identifies the cells from the background	Popular in cell image analysis	Issue of under segmentation and over-segmentation depending on cell complexity
Marker Controlled (MC) watershed	SM watershed separate the overlapped cell	This method does not achieve great accuracy but extracted features supports cell classification	
Split and Merge (SM) watershed			
Active contour segmentation methods	DCAC employs initial contour to the previously delineated nuclear boundary by considering intensity information	Works well for nuclei having a clear nuclear boundary	Need of initial contour, proper parameters, and long runtime
Dual channel active contour [9] (DCAC)	The edge-based model uses local edge information to fit the boundaries of the approximated shape	Autonomous and self-adapting method of finding object boundary	Difficult to segment the nearest objects
Edge based active contour [10]			When image size is too large, this method works slowly

pool the multilevel features from the hierarchical architecture then fuse it together that improves the performance of the model. Better visualization of object boundary by proposing contour-based architecture make the model efficient. In a segmentation process, pooling operation losses some relevant information; to address this problem, Fisher Yu et al. [13] merge the multiscale information by aggregating the output coming from different dilation rates to regain the resolution. This architecture is the best fit for dense prediction due to the expansion of the receptive field. Semantic segmentation of the road and indoor called SegNet [14] has encoder, decoder, and pixel-wise classification stages. To upsample the low-resolution feature map, utilization of max-pooling indices in the decoder side provides exact boundary information which is very important in any road and indoor understanding. [15] proposed a fully convolutional model which is designed for semantic segmentation. Since their model is inspired by AlexNet which is a classification model and is extended for the segmentation task because it does classification and localization both. [16]’s method is quick and capable to process large batches of data in a sensible amount of time. Their deep learning architecture effectively combines the concept of ResNet and UNet. Recognition of close and overlapped nuclei was the major concern. This particular issue was addressed in [17] by predicting an eroded version

of the annotated nuclei. They formulated a global loss function to solve the issue of segmentation of overlapped nuclei. To increase segmentation accuracy a meaningful modification in UNet architecture by implementing an attention gate in [18], which is able to merge only relevant features before concatenation. The best part of this work is that without much computational complexity the attention module can be integrated with any other segmentation model. Different dilation rate is mostly equivalent to different sparse kernels without the extra computational overhead. The idea is effective in [19] to enlarge the field of view. Convolution strategies have a big impact on the efficiency of the system.

The idea of dimension-wise convolution has been implemented by [26] to improve the performance for segmentation and classification tasks. To extract the semantic features of nuclei at different stages [22] incorporated an improved version of atrous spatial pyramid pooling in the encoder–decoder network which used concave point detection method segment touching nuclei and achieved better performance. ASPPU module in combination with binary cross-entropy loss and dice loss better handle the class imbalance problems. One of the recent concepts by [23] is UNet++ for medical image segmentation. UNet++ modify the skip connection method to provide flexibility in the fusion of multilayered features in the decoder path of the network and also their network shares the

common encoder for varying depths. Segmentation of liver histopathology images having thousands of nuclei and many of them are overlapped. [24] incorporated residual and attention mechanism in encoder–decoder architecture. Residual block to regain the meaningful information, attention block for correct localization of objects in the decoder section, and bottleneck layer for the maximum extraction of deeper layer features. From the above existing work, it is clear that segmentation of nuclei from histopathological images is still a challenging issue. A brief comparison of existing deep learning approaches is presented in Table 2. The novelty of the proposed model lies in the manner in which we extended the idea of UNet [11] and ASPPU-Net [22] such that the proposed network is efficient and capable to extract intermediate features.

## Proposed architecture

In deep learning framework segmentation strategy as in [11,18,24] involves encoder path where model learns the ‘what’ information or context in the image. In decoder path, the model learns the ‘Where’ information or location in the image. The encoder path consists of repeated application of standard convolution layer and an additional parallel path to the main network through a residual connection to minimize the loss in pooling operation. Each stage of the downsampling path has a  $(2 \times 2)$  max-pooling layer which effectively reduces the spatial size of the image. Similar level features are concatenated with upsampling layers. In the upsampling path, the features having different shapes and size has cropped and merged into the next layer. Our wide network used a powerful decoder by aggregating the similar level spatial features and collecting maximum possible residual information. At the final stage, a  $(1 \times 1)$  convolution is used to map the size  $(512 \times 512 \times 16)$  to  $(512 \times 512 \times 1)$ . To improve the performance of the network shown in Fig. 1, we introduce an additional information retrieval module to extract a more relevant feature used in [22] called ASPPU bottleneck path where a CNN uses multiple dilation rate. Dilation rate is an additional parameter that differed from the resultant feature map to visualize a larger area. Different dilation rates are applied to the same layer and results are then concatenated, which allows the model to take advantage of multiscale feature extraction. This yields very good results because hierarchical information with varying sizes can be identified in the same layer.

## Standard convolution layer

Convolution layers consist of a set of learnable parameters. Our input image has three color channels (RGB) and has a dimension of  $512 \times 512 \times 3$ , on which we are applying a filter

of size  $(3 \times 3)$ . If we talk about standard 2D convolution and if we have  $x(p,q)$  is the input feature map and  $h(p,q)$  is the filter kernel, then  $y(p,q)$  is the output of 2D convolution expressed in Eq. (1).

$$y(p, q) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} x(j, k)h(p - j, q - k). \quad (1)$$

When we apply a kernel of  $K \times K$  on the image of  $N \times N$  with padding  $P$  and stride  $S$ , we will get the image size as in Eq. (2).

$$N \times N \rightarrow \left( \frac{N - K + 2P}{S} + 1 \right) \left( \frac{N - K + 2P}{S} + 1 \right). \quad (2)$$

Without padding and single-step, stride, the relation between the input image and the output image is in Eq. (3).

$$N \times N \rightarrow (N - K + 1) (N - K + 1). \quad (3)$$

## High-resolution layer

Proposed architecture addressed the problem of degradation of information in deeper network by introducing deep and wide residual network which is easier to train and optimize. The residual connection is realized by creating an additional path parallel to the main encoder–decoder path of the network to restore the flow of information in deep network. Instead of deep and thin encoder and the decoder path our network has wide and deep path that effectively leverage the strength of residual learning as well as encoder–decoder architecture. Proposed architecture has a high-resolution encoder in Fig. 2a, ASPP bottleneck path in Fig. 2b for multilevel feature extraction and an effective decoder shown in Fig. 2c.

## Activation function

Any linear function has less capability to learn any complex input. Since the outcome of the convolution layer is based on linear operation, a nonlinear operation is needed to map complex input to a meaningful output. The nonlinear activation function allows the network to learn complex data, compute and learn almost any function, and provide accurate predictions. It also helps the model to generalize or adapt to a variety of data and to differentiate the output. Rectified Linear Unit (ReLU) is the most popular activation function in deep learning models. ReLU takes care of several problems faced by the sigmoid and the tanh activation function, and it also avoids and rectifies the vanishing gradient problems. The ReLU activation function simply has a value of zero, if it receives any negative input, but for any positive value, it

**Table 2** Deep learning image segmentation methods

References and dataset	Application	Advantages	Limitations
Olaf Ronneberger et al. (2015) [11]. PhC-U373, DIC-HeLa (ISBI cell tracking images)	Cell segmentation and various biomedical image segmentation tasks	A benchmark model for biomedical image segmentation  Can be easily scaled	For deeper model, learning gets slow  Ground truth mask required High GPU required for larger images
Hao Chen et al. (2016) [12]. Warwick-Qu Gland Dataset	Gland segmentation from histopathology data	Network provides clear contour of gland images  A unified deep learning framework	Not advantageous in the case of large-scale histopathology data
Fisher Yu et al. (2016) [13]. Pascal VOC 2012 segmentation Dataset	Object segmentation and Class segmentation	Systematically aggregated the semantic features of different stages  Designed network is useful for dense prediction	This network is not an end-to-end framework  Uses pre-trained weights
Vijay Badrinarayanan et al. (2017) [14]. CamVid road scene Dataset & SUN RGB-D Indoor Scene Dataset	Road and Indoor segmentation	Improved boundary delineation and less memory requirement	Semantic pixel-wise labeling is a tedious task
Evan Shelhamer et al. (2017) [15]. Pascal VOC, SHIFT FLOW & NYUDv2 Segmentation Dataset	Semantic segmentation of objects like Person, Animal, Vehicle, etc.	End-to-end trained network	Segmented boundaries are not fine enough
Yves-Remi et al. (2018) [16]. Glas Challenge Dataset	Gland Segmentation from histopathology images	Augmented the training data by varying the color and intensity properties  Effectively Combines the concept of ResNet and UNet	The visibility of segmented nuclei is a major concern of this approach
Peter Naylor et al. (2019) [17]. Triple Negative Breast Cancer (TNBC) & MoNuSeg Dataset	Segmentation of Nuclei from Multi-Organ histopathology images	Focused to separate the clumped nuclei  Introduced the intra-nuclear distance concept for nuclei segmentation	The problem of segmentation of close nuclei is partially solved but not completely
Jo Schlemper et al. (2019) [18]. 3D CT Abdominal Dataset & 2D Fetal ultrasound Image Dataset	Medical image segmentation and classification	It suppresses the irrelevant features while highlights the meaningful features  Attention module can be attached in most of the CNN-based segmentation and classification framework	It increases the training time of the model particularly for a large sequence of data

**Table 2** continued

References and dataset	Application	Advantages	Limitations
Xipeng Pan et al. (2019) [20]. Triple Negative Breast Cancer (TNBC) & MoNuSeg Dataset	Segmentation of Nuclei from Multi-Organ histopathology images	Used combined loss function that better train the model for pathology dataset  Depthwise separable convolution is used to control the trainable parameter	Segmented cell boundaries differ from ground truth images
Simon Graham et al. (2019) [21]. GlaS Challenge Dataset	Lumen and Gland Segmentation in Colon histopathology Images	This method was the first to apply rotational symmetry in encoder–decoder architecture  The additional rotated kernel in G-convolution better extracts the discriminative features	Improvement is marginal
Tao Wan et al. (2020) [22]. Four H&E Stained Dataset(TNBC, MoNuSeg, 80 WSI of Lung histopathology images, 58 invasive Breast carcinoma)	Nuclei segmentation from Multi-Organ histopathology images	Captures multi-layer features of nuclei by integrating ASPPU module and concave point detection  Joint loss function better handles the class imbalance problems	The applied post-processing technique is not providing promising results
Zongwei Zhou et al. (2020) [23]. ISBI-2012, Cell-CT, BraTS-2013, DBS-2018, MICCAI-2017, LIDC-IDRI	Medical image segmentation (Microscopy, CT, MRI)	Provides flexibility in the fusion of multilayered features  Network shares the common encoder for varying depths  An improvement over the fixed-depth U-Net	The architecture becomes nested  Extended decoders and redesigned skip connections result in longer training time
Shyam Lal et al. (2021) [24]. KMC Liver Dataset & MoNuSeg Dataset	Segmentation of Nuclei from Liver and Multi-Organ histopathology images	Fully automated deep learning framework  For the nuclei segmentation task, this architecture is faster than UNet [11], Dist [17], and Attention UNet [18]	Morphology of detected nuclei differs from the original mask and some nuclei near the boundary region are missing
Mahendra Khened et al. (2021) [25]. H&E Stained- Breast cancer (CAMELYON), Colon cancer (Digest path), Liver cancer (PAIP)	H&E Stained segmentation of pathology images	Processed the entire whole slide image  Utilized multiple fully connected network architectures  Generated the uncertainty maps, which is helpful in fine-tuning the model	Lack of generalizability in the proposed method

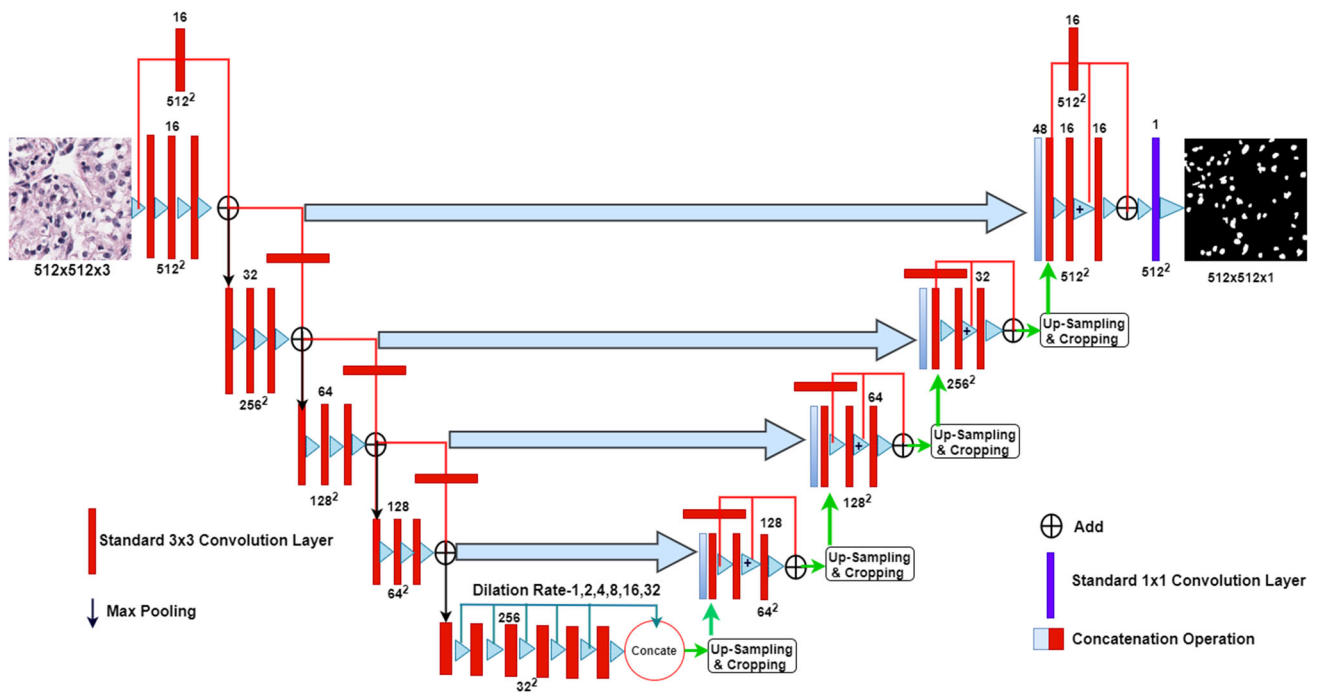


Fig. 1 Proposed high-resolution deep transferred ASPPU-Net

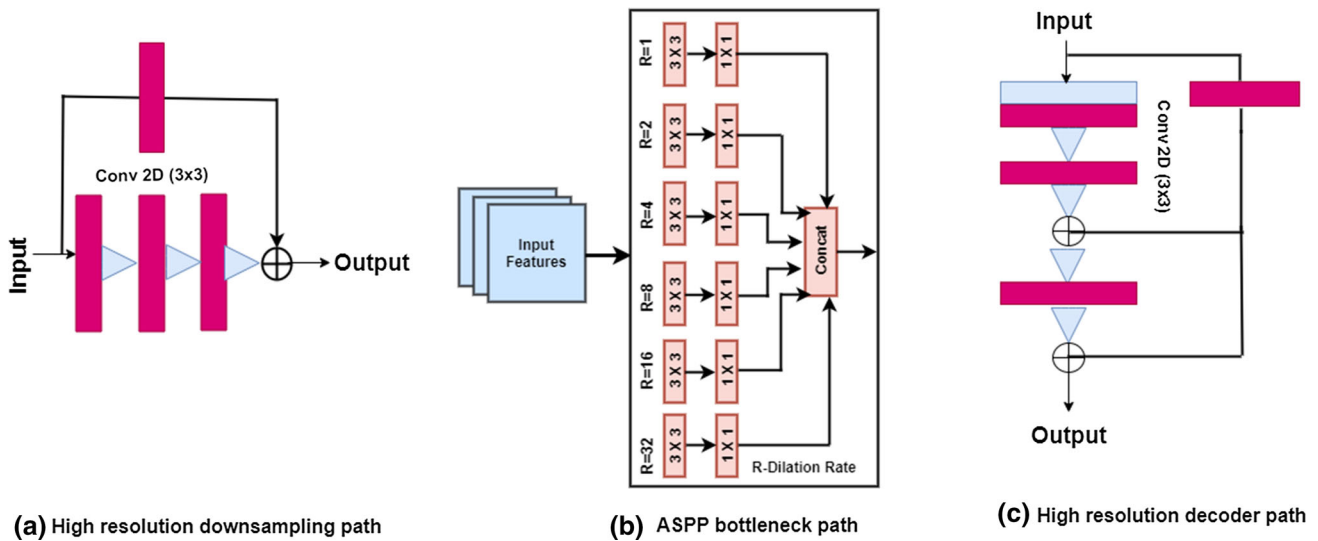


Fig. 2 High-resolution encoder path, ASPP bottleneck path, and decoder path

returns that value like a linear function. It is computationally economical compared to sigmoid and tanh. Mathematically, ReLU activation function and its derivative are expressed by Eqs. (4) and (5).

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases} \quad (4)$$

$$f'(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases} \quad (5)$$

**Pooling layer**

Pooling is a concept that makes our model, location invariant, scale-invariant, rotation invariant. It acts as an additional layer. Whatever we have a higher value within the kernel dimension, where higher value corresponds to object region, only that area will reflect in the higher layer which makes detection task easy. By applying a kernel of size  $K = (2 \times 2)$  and stride size  $S = 2$  on  $(4 \times 4)$  two-dimensional image, the resultant feature size is shown in Fig. 3.



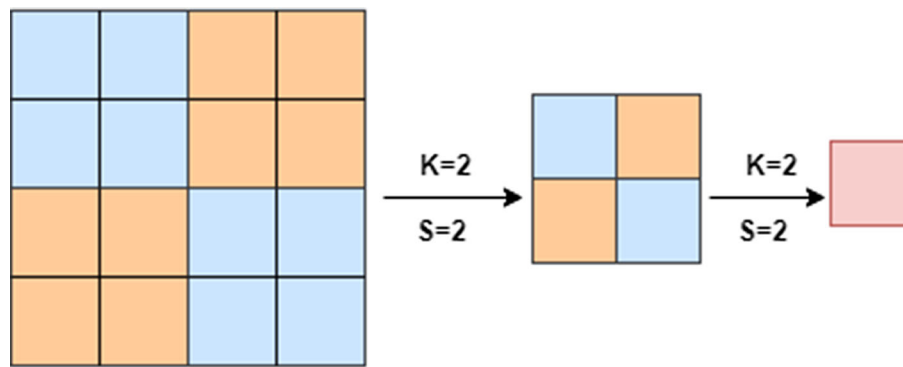


Fig. 3 Max-pooling with kernel  $K = 2$  and stride  $S = 2$

**Batch normalization**

Due to lots of operations happening between layers, if the input changes slightly in the deep network, this can lead to large changes at later layers. The distribution of each layer’s input changes during training. As we train the network weight keeps changing. If the weight changes randomly in successive iterations, it becomes difficult for the network to adjust each layer’s input. Batch normalization [27] makes sure the distribution does not change too much. During the training procedure gradient of the loss, plays an important role in the convergence. Visualization of internal covariate shift in the deeper network can be seen in Fig. 4. Batch normalization we can think of as an additional layer and it works well in the very deep network. It has the following advantages (a) faster convergence, (b) works as a regularizer, (c) avoids internal covariate shift, and (d) we can train a deeper network.

Let  $T_1, T_2$  are two transformations at a particular layer  $L$  shown in Eq. (6). Two layers are characterized by weights  $W_1$  and  $W_2$ . If  $W_1$  changes,  $T_1$  will change, and input to  $T_2$  will change. If  $W_2$  changes  $L$  will change. If these changes are random and large then there is the problem of convergence in deep neural networks.

$$L = T_2(T_1(U, W_1), W_2) \tag{6}$$

$$\text{Batch Norm } (y_o)^{(m)} = \beta_1 \left( \frac{(y_o)^{(m)} - \mu(y_o)}{\sigma(y_o)} \right) + \beta_2 \tag{7}$$

Batch normalization normalizes each activation independently by controlling the mean and standard deviation of the layer’s output. The process expressed in Eq. (7), where  $y_o^m$  is the value of the output  $y_o$  on the  $m$ th input of a batch and  $\beta_1, \beta_2$  are trainable parameters.

**Internal covariate shift**

Santurkar et al. [28] analyzed the effect of internal covariate shift, here the idea is the constant changes in the layer’s input

distribution are beneficial for fast training. Their experiment also suggests that for larger learning rate training may not converge without batch normalization. Impacts of internal covariate shift on optimization, where it measures the effect at the  $k^{th}$  layer of  $n$  layer network having parameters  $W_{1:n}$  and  $W'_{1:n}$  as  $\|\nabla_{W_k} \mathcal{L}(W_{1:n}) - \nabla_{W_k} \mathcal{L}(W'_{1:k-1}, W_{k:n})\|$ . It follow the properties of convex function.

If  $D$  is a convex subset of  $\mathbb{R}^n$  and  $\nabla_X f, X, Y \in \mathbb{R}^n$  Let  $f : D \mapsto \mathbb{R}$  is a convex function then

1.  $f$  is  $L$  Lipschitz if  $|f(Y) - f(X)| \leq L \|Y - X\| \quad \forall X, Y \in D$
2.  $f$  is  $\beta$ -smooth if its gradient is  $\beta$ -Lipschitz, i.e.,  $\|\nabla_X f - \nabla_Y f\| \leq \beta \|X - Y\| \quad \forall X, Y \in D.$

**Dilated convolution layer**

The mathematical expression of a 2-D dilated convolution is given by Eq. (8).

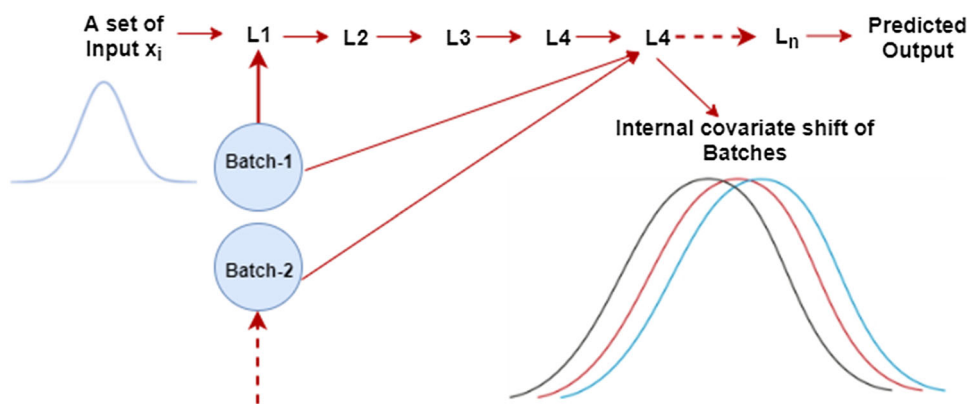
$$y(p, q) = \sum_{j=1}^P \sum_{i=1}^Q x(p + r * i, q + r * j) f(i, j) \tag{8}$$

$y(p, q)$  is the output of dilated convolution,  $x(p, q)$  is the input feature map,  $f(i, j)$  is the filter kernel and,  $r$  is the rate ( $r > 1$ ) with length and width of  $P$  and  $Q$ , respectively.

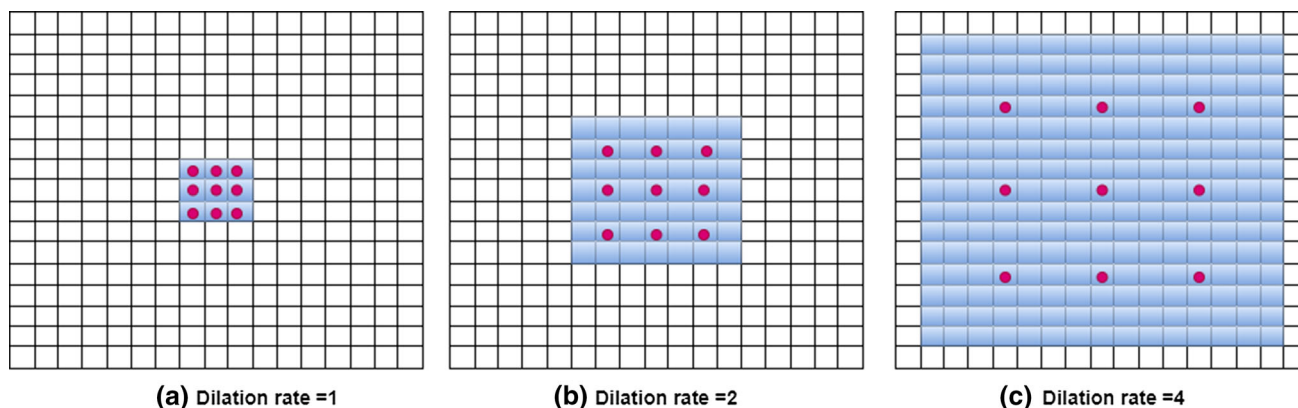
Dilated convolution or atrous convolution is a simple way to increase the field of view by placing spaces between the elements of the kernel. The rate of dilation is controlled by the parameter  $r$ . Dilation rate  $r$ , insert  $(r - 1)$  spaces between the elements of the kernel.  $r$  equal to 1 corresponds no space, called standard convolution. A kernel of size  $K$  dilated by a factor  $r$  has an effective size in Eq. (9).

$$K' = K + (K - 1)(r - 1) \tag{9}$$

Dilated convolution is similar to standard convolution applied to the input with different gaps.



**Fig. 4** Internal covariate shift



**Fig. 5** Visualization of the receptive field with multiple dilation rates

Visualization of the receptive field with different dilation rates is shown in Fig. 5. If we applied dilation rate 1 to the input 2D image in Fig. 5a, it is similar to standard convolution with kernel size  $3 \times 3$ . Dilation rate 2 produces a receptive field of  $7 \times 7$  by skipping one pixel shown in Fig. 5b. Dilation rate 4 produces a receptive field of  $15 \times 15$  by skipping three pixels shown in Fig. 5c. The motive of this work is to extract more relevant features by applying different dilation rates in the same layer. Dilated convolution allows us to expand our filter strides with different rates, so this convolution becomes dilation convolution. In this architecture, the maximum relevant information is retrieved by applying three different dilation rates after each pooling layer. Dilated convolution [29] is a generalization of a standard convolution that allow us to control the resolution of features computed by deep CNN in order to capture multi-level feature.

## Implementation and training

This experiment is implemented on the latest version of python with Keras and TensorFlow framework. Google GPU and Colab notebook were the resources for conducting the experiments. We have used three histopathological

datasets from the literature [17,30,31]. (a) Kidney dataset: This dataset of size  $400 \times 400$  pixels is used in [29], and consists of 730 H&E renal cell carcinoma (RCC) histology images created from 10 WSIs of TCGA. (b) Triple-Negative Breast Cancer (TNBC) dataset: TNBC dataset [30] consists of 33 H&E stained breast tissue of dimension  $512 \times 512$  pixels, and these tissues are collected from seven different patients. (c) MoNuSeg dataset: This dataset is first used in [31] and composed of 30 H&E stained histology images of size  $1000 \times 1000$  pixels of seven organs. We partitioned the obtained data into training, validation, and test. After the creation of patches, we considered 80% of the total images for the training of the model and the rest 20% for validation and testing. We had applied some data augmentation techniques such as horizontal flip and vertical flip in the training sample of the TNBC, and MoNuSeg dataset. Adam [32], is the optimization method and binary cross-entropy used in [17] is the loss function used in this study. The reported results of all deep learning models are the averages of three trials conducted independently. We had calculated the final results based on an average of three trials by initializing random weights in each trial. The final quality metrics were measured as the average quality metrics of all images in the test dataset.

**Table 3** Experimental performance comparison of different architectures with three datasets

Model	Kidney		TNBC		MoNuSeg	
	F1	AJI	F1	AJI	F1	AJI
UNet(2015)	0.8537	0.7489	0.7324	0.6559	0.7248	0.6029
SegNet(2017)	0.8972	0.8304	0.7685	0.6434	0.7879	0.6514
Attention UNet(2019)	0.9135	0.8590	0.7216	0.6194	0.7735	0.6315
DIST(2019)	0.8992	0.8272	0.7516	0.6727	0.7795	0.6467
ASPPU-Net(2020)	0.9052	0.8293	0.7781	0.6378	0.8087	0.6803
Proposed model	0.9684	0.9394	0.8419	0.7282	0.8344	0.7169

**Table 4** Network computational complexity in millions

Model	Parameters (millions)	FLOPs (millions)
UNet	31.3	62.7
SegNet	18.8	47
Attention UNet	31.9	63.7
Dist	7.7	15.5
ASPPU-Net	4.3	8.6
Proposed	8.7	21.6

**Performance metrics**

*Jaccard coefficient* It is a commonly used method for the measurement of overlap of two sets. It is a measure of similarity or dissimilarity between binary data. Jaccard coefficient measurement is the best method to evaluate the performance in the case of nuclei segmentation from histopathology images in a deep learning framework.

$$Jaccard(A, B) = \frac{\text{Number of items in } A \cap B}{\text{Number of items in } A \cup B}$$

If we take the Jaccard coefficient of a set with itself ratio will be one and the Jaccard coefficient will be one.

$$Jaccard(A, B) = 1.$$

If two sets are disjoint and have no members in common, then Jaccard coefficient will be zero.

$$Jaccard(A, B) = 0.$$

If two sets are not the same size, then the Jaccard coefficient will always assign a number between zero and one.

$$0 \leq Jaccard(A, B) \leq 1.$$

Jaccard distance/ Jaccard loss = (1 – Jaccard coefficient).

*Aggregated Jaccard index (AJI)* The concept of AJI was proposed in [31]. AJI compute and measure the performance

of segmentation better than the global Jaccard index by incorporating the concept of the connected components. AJI compute and measure the performance of segmentation better than the global Jaccard index by incorporating the concept of the connected components.

$$AJI = \frac{\sum_{k=1}^M |G_k \cap P_S^k|}{\sum_{k=1}^M |G_k \cup P_S^k| + \sum_{R \in U} |P_R|}$$

where  $G_k$  is  $k$ th nuclei of ground truth having M nuclei.  $P_S^k$  is  $S$ th Connected component in prediction result having the highest Jaccard index with ground truth. Each index of  $S$  cannot be utilized more than once.  $U$  represents a set of the connected components in the prediction result without corresponding ground truth.  $P_R$  represents the ground truth and intersection that are not in the associated space. AJI is a connected component-based method that is the improved version of the pixel-based global Jaccard index. A higher AJI value indicates a better segmentation method.

*Accuracy* Accuracy is a good measure only when we have symmetric dataset. For data being symmetric, Values of false-positive should be almost the same as false negatives.

$$Accuracy = \frac{[TP + TN]}{[Total \ predicted \ pixels]}$$

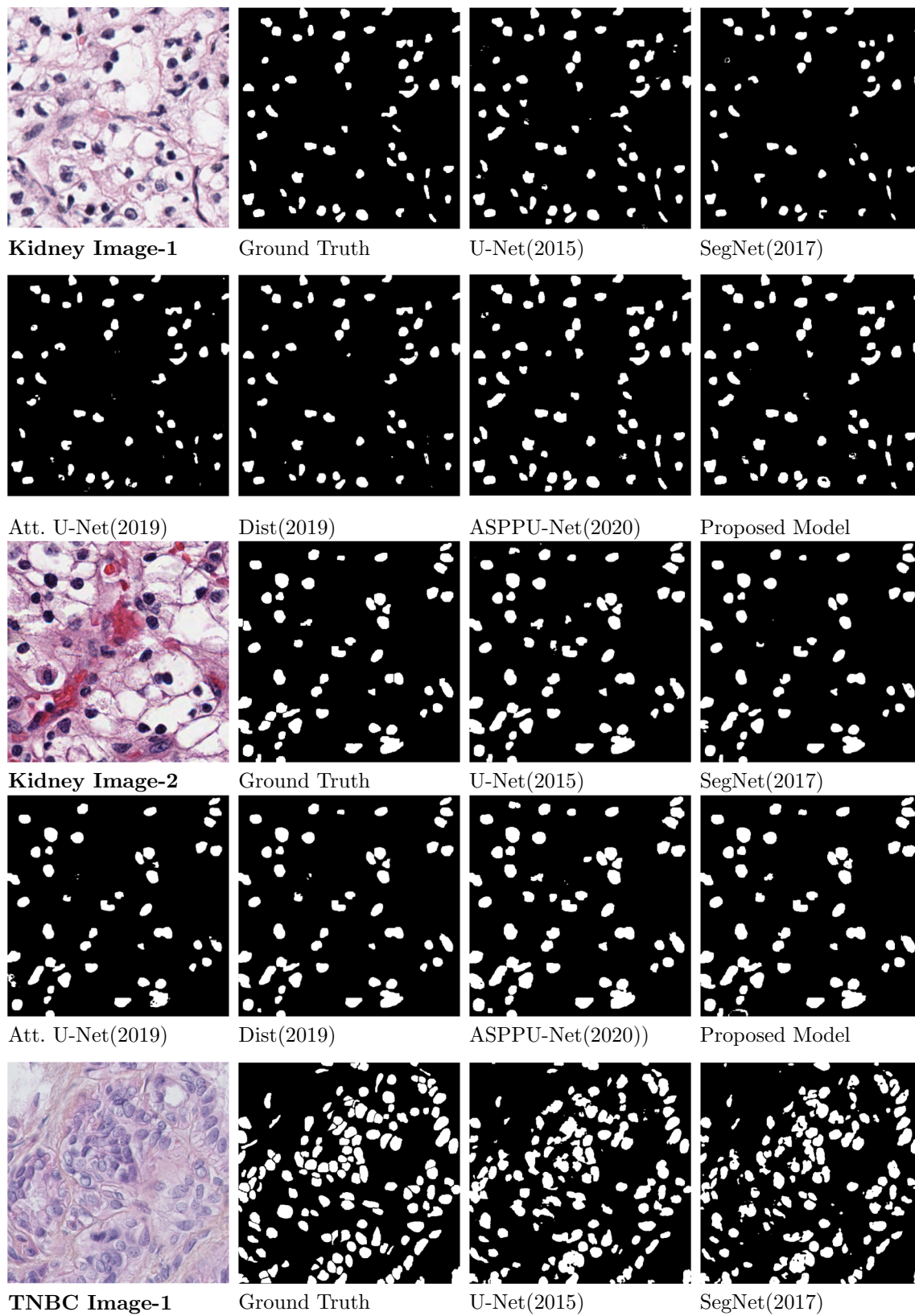
*Precision* It tells out of total predicted positive observation how much sample is true positive.

$$Precision = \frac{TP}{[Total \ positive \ prediction]}$$

*Recall/Sensitivity* It calculates the ratio of correctly predicted positive observation of the observation in actual class.

$$Recall/Sensitivity = \frac{TP}{[TP + FN]}$$

*F1 Score* F1 score has the effect of both precision and recall by calculating the harmonic mean between precision and recall. F1 Score is the best method to measure that how much information is retrieved used by [33]. The range of the F1



**Fig. 6** Row-wise visual segmentation comparison of different models on three datasets

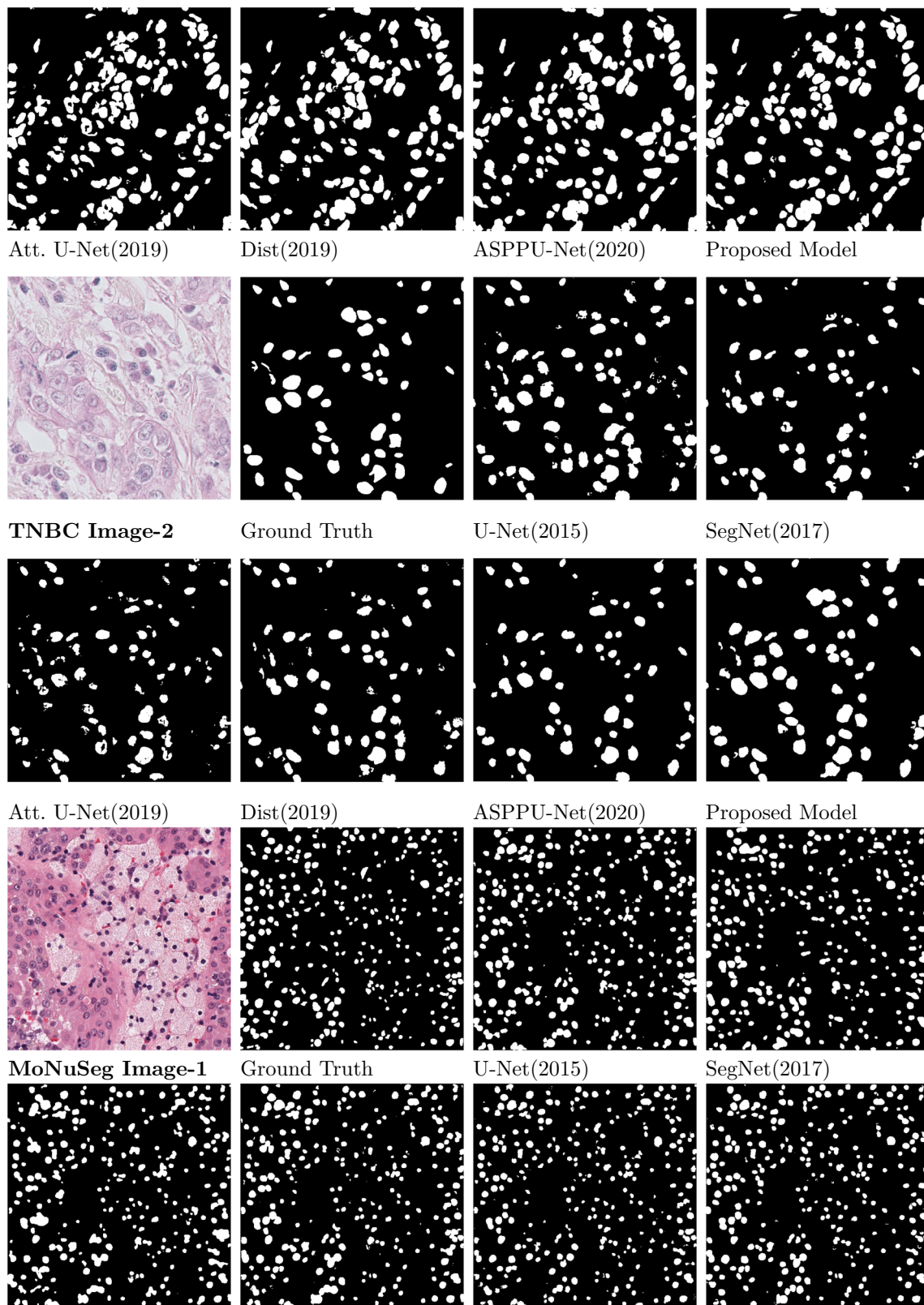


Fig. 6 continued

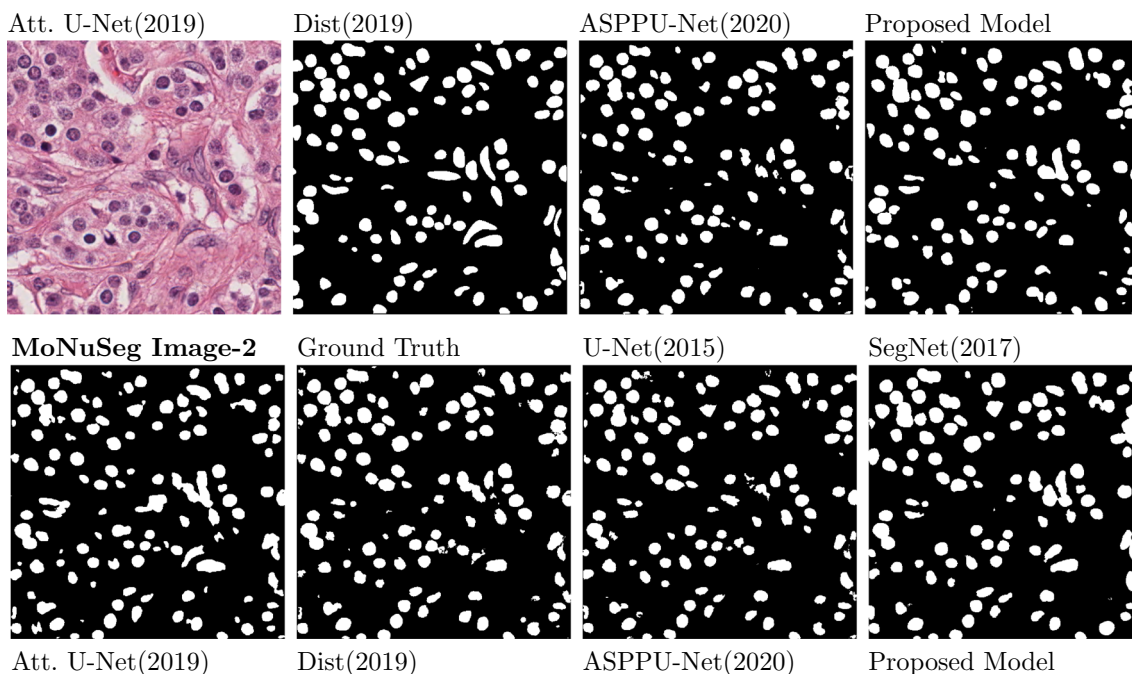


Fig. 6 continued

Score is [0, 1]. The greater the F1 score, the better is the performance of our model.

$$\text{F1 Score} = 2 \times \frac{[\text{Recall} \times \text{Precision}]}{[\text{Recall} + \text{Precision}]}$$

## Results

### Comparison of methods

Table 3 shows a comparison of the proposed architecture with five other segmentation architectures on three datasets. Performance measurement in terms of F1 Score, AJI score, the total number of trainable parameters that describe the training time and complexity and floating point operations per second (FLOPs). The value of FLOPs describes the computing power of given hardware like GPU, the smaller the value, the faster is the computing ability. Here, our F1 Score and AJI score are the averages of all images of the test set in the dataset. Table 4 shows the comparison of the total number of trainable parameters and FLOPs with other segmentation architectures.

### Performance comparison of different architectures

Visual segmentation comparisons of predicted images by different models on the Kidney, TNBC, and MoNuSeg datasets are shown in Fig. 6. Visual results in Fig. 6 are the comparison of predicted images of five state-of-the-art models

with the proposed model of two images of each of the three mentioned datasets. For the purpose of discussion on segmentation accuracy, we are taking one sample test image from the kidney dataset shown in row-1 of Fig. 6. Their corresponding ground truth image has 57 annotated nuclei. U-Net [11], detected 53 nuclei clearly, two nuclei are not detected and two are in clustered form. Four additional ducts are detected in U-Net which are not desirable and not present in the ground truth. SegNet [14], also not detected any additional things, but out of 57 nuclei only 47 nuclei detected accurately, three are detected partially and seven nuclei are not detected. The best part of SegNet architecture is that it separates the overlapped nuclei. Attention U-Net [18] is an extended version of the original U-Net that detected 49 nuclei clearly, two nuclei are in overlapped form, five nuclei are partially detected, one nucleus not detected and five additional things found that are not required. Dist [17], architecture detected 49 nuclei almost similar to ground truth, three nuclei not able to detect, three partially detected, and two are in clustered form. ASPPU-Net [22] identifies 7 additional shapes, 2 nuclei are in clustered form out of 57 nuclei.

The best part of our proposed model is out of 57 nuclei, 55 nuclei are clearly identified, no undesirable things are detected, the morphology of the detected nuclei are exactly similar to the ground truth image and it also reflects from performance metrics shown in tabular form. Results indicated that our architecture is able to retrieve more information compared to others. Training and validation accuracy, training, and validation loss curve on three datasets are shown in

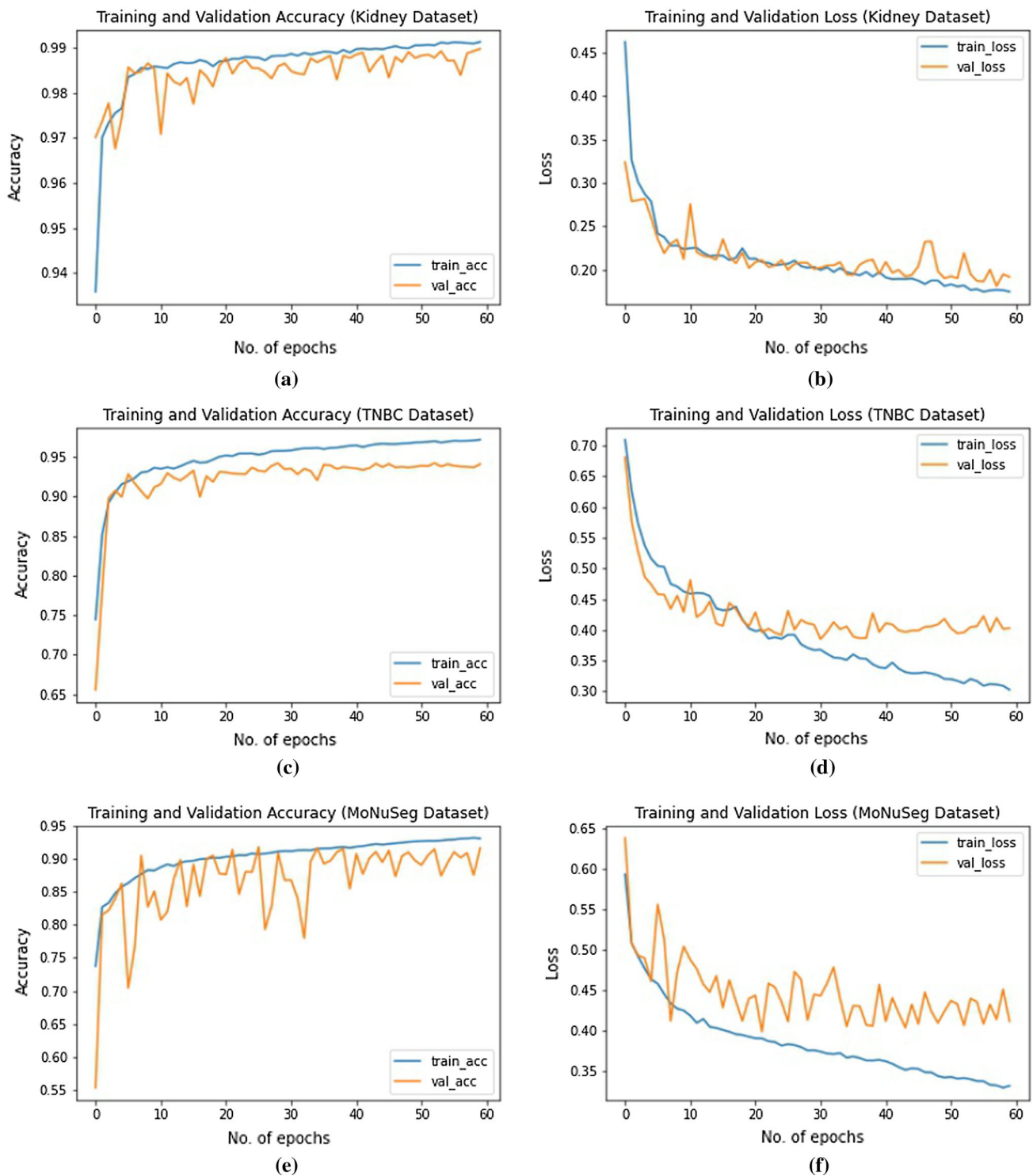


Fig. 7 Training accuracy, validation accuracy, training loss, validation loss curve on three datasets

Fig. 7. These plots indicated that the proposed model better learns the parameters for the prediction on each of mentioned datasets.

### Conclusion

Automatic segmentation of H&E stained cell nuclei from histopathology image is a pre-requisites in digital pathology. In this paper, a CNN-based architecture called high-

resolution deep transferred ASPPU-Net was proposed that addressed automatic nuclei segmentation of histopathology images having a widely varied spectrum with a large number of artifacts. The implemented networks effectively leverage the strength of residual learning as well as encoder–decoder architecture by incorporating wide and deep network paths that strengthen the intermediate features. Promising results were obtained due to the effective use of the wide and deep network with ASPP at the bottleneck layer. To prove the worthiness of the proposed architecture, we have used the most preferred performance matrices F1 score and AJI score by performing experiments on three different publically available datasets. The proposed model achieved a considerable margin in terms of F1 score and AJI score over state-of-the-art models and works effectively for three histopathological datasets.

Although the proposed model produced excellent results, the segmentation of overlapped nuclei is still a challenge for some histopathology images and reported results are still sub-optimal for clinical use. Furthermore, vanished and blurry boundaries of detected nuclei are another issue. The problem of over-segmentation and under-segmentation for nuclei associated with complex histopathology images has not been solved completely. These issues will be the focus of our future work.

**Acknowledgements** The authors would like to thank the Editor-in-Chief and anonymous Reviewers for their constructive comments, which improved the quality of the manuscript.

**Funding** This research work was supported in part by the Science Engineering and Research Board, Department of Science and Technology, Govt. of India under Grant No. EEG/2018/000323, 2019.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** We have used three publically available datasets.

## References

1. Washington MK, Berlin J, Branton P, Burgart LJ, Carter DK, Fitzgibbons PL, Halling K, Frankel W, Jessup J, Kakar S, Minsky B, Nakhleh R, Compton CC (2009) Protocol for the examination of specimens from patients with primary carcinoma of the colon and rectum. *Arch Pathol Lab Med* 133:1539–1551
2. Gonzalez RC, Woods RE (2006) *Digital image processing*, 3rd edn. Prentice Hall, New York
3. Win K Y, Choomchuy S, Hamamoto K (2017) K mean clustering based automated segmentation of overlapping cell nuclei in pleural effusion cytology images. In: International conference on advanced technologies for communications (ATC), pp 265–269. <https://doi.org/10.1109/ATC.2017.8167630>
4. Dhanachandra N, Chanu YJ (2020) An image segmentation approach based on fuzzy c-means and dynamic particle swarm optimization algorithm. *Multimed Tools Appl* 79:18839–18858. <https://doi.org/10.1007/s11042-020-08699-8>
5. Stutz D, Hermans A, Leibe B (2018) Superpixels: an evaluation of the state-of-the-art. *Comput Vis Image Understanding* 166:1–27. <https://doi.org/10.1016/j.cviu.2017.03.007>
6. Albayrak A, Bilgin G (2019) Automatic cell segmentation in histopathological images via two-staged superpixel-based algorithms. *Med Biol Eng Comput* 57(3):653–665
7. Cousty J, Bertrand G, Najman L, Couprie M (2010) Watershed cuts: thinnings, shortest path forests, and topological watersheds. *IEEE Trans Pattern Anal Mach Intell* 32(5):925–939. <https://doi.org/10.1109/TPAMI.2009.71>
8. Gamarra M, Zurek E, Escalante HJ, Hurtado L, San-Juan-Vergara H (2019) Split and merge watershed: a two-step method for cell segmentation in fluorescence microscopy images. *Biomed Signal Process Control*. <https://doi.org/10.1016/j.bspc.2019.101575>
9. Song T, Sanchez V, Eidaly H, Rajpoot NM (2017) Dual-channel active contour model for megakaryocytic cell segmentation in bone marrow trephine histology images. *IEEE Trans Biomed Eng* 64(12):2913–2923
10. Ciecholewski M (2016) An edge-based active contour model using an inflation/deflation force with a damping coefficient. *Expert Syst Appl* 44:22–36
11. Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. In: *Proceedings MICCAI*. Springer, Munich, pp 234–241
12. Chen H, Qi X, Yu L, Heng PA (2016) DCAN: deep contour-aware networks for accurate gland segmentation. *Computer Vision and Pattern Recognition*. [arXiv:1604.02677v1](https://arxiv.org/abs/1604.02677v1) [cs.CV]
13. Yu F, Koltun V (2016) Multi-scale context aggregation by dilated convolutions. Published as a conference paper at ICLR, San Juan, PR, USA, pp 1–13. <https://arxiv.org/pdf/1511.07122.pdf>
14. Badrinarayanan V, Kendall A, Cipolla R (2017) SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell* 39(12):2481–2495
15. Shelhamer E, Long J, Darrell T (2017) Fully convolutional networks for semantic segmentation. *IEEE Trans Pattern Anal Mach Intell* 39(4):640–651
16. Remi Y, Eycke V, Balsat C, Verset L, Debeir O, Salmon I, Decaestecker C (2018) Segmentation of glandular epithelium in colorectal tumours to automatically compartmentalise ihc biomarker quantification: a deep learning approach. *Med Image Anal* 49:35–45
17. Naylor P, Lae M, Reyat F, Walter T (2019) Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE Trans Med Imaging* 38(2):448–459
18. Schlemper J, Oktay O, Schaap M, Heinrich M, Kainz B, Glocker B, Rueckert D (2019) Attention gated networks: learning to leverage salient regions in medical images. *Med Image Anal* 53:197–207
19. Kim J, Heo YS (2019) Efficient semantic segmentation using spatio-channel dilated convolutions. *IEEE Access* 7:154239–154252
20. 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
21. Graham S, Epstein D, Rajpoot N (2019) Rota-Net: rotation equivariant network for simultaneous gland and lumen segmentation in colon histology images. *European Congress on Digital Pathology*, Springer Cham, pp 109–116
22. Wan T, Zhao L, Feng H, Li D, Tong C, Qin Z (2020) Robust nuclei segmentation in histopathology using ASPPU-Net and boundary refinement. *Neurocomputing* 408:144–156
23. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J (2020) Unet++: redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Trans Med Imaging* 39(6):1856–1867



24. Lal S, Das D, Alabhya K, Kanfade A, Kumar A, Kini J (2021) NucleiSegNet: robust deep learning architecture for the nuclei segmentation of liver cancer histopathology images. *Comput Biol Med* 128:104075
25. Khened M, Kori A, Rajkumar H, Krishnamurthi G, Srinivasan B (2021) A generalized deep learning framework for whole slide image segmentation and analysis. *Sci Rep* 11:11579
26. Mehta S, Hajishirzi H, Rastegari M (2019) Dicenet: dimension-wise convolutions for efficient networks. *Computer Vision and Pattern Recognition*. [arXiv:1906.03516](https://arxiv.org/abs/1906.03516)
27. Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. *Machine Learning*. [arXiv:1502.03167](https://arxiv.org/abs/1502.03167)
28. Santurkar S, Tsipras D, Ilyas A, Madry A (2019) How does batch normalization help optimization. *Neural and Evolutionary Computing*. [arXiv:1805.11604](https://arxiv.org/abs/1805.11604)
29. Chen LC, Papandreou G, Schroff F, Adam H (2017) Rethinking atrous convolution for semantic image segmentation. *Computer Vision and Pattern Recognition*. [arXiv:1706.05587v3](https://arxiv.org/abs/1706.05587v3)
30. Irshad H, Kouhsari LM, Waltz G, Bucur O, Nowak JA, Dong F, Knoblauch NW, Beck AH (2015) Crowdsourcing image annotation for nucleus detection and segmentation in computational pathology: evaluating experts, automated methods, and the crowd. In: *Pacific Symposium on Biocomputing (PSB)*. <https://doi.org/10.13140/2.1.4067.0721>
31. 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 Trans Med Imaging* 36(7):1550–1560
32. Kingma DP, Ba J (2015) Adam: a method for stochastic optimization. In: *International conference on learning representations*, p 9. [arXiv:1412.6980v9](https://arxiv.org/abs/1412.6980v9)
33. Chanchal AK, Kumar A, Lal S, Kini J (2021) Efficient and robust deep learning architecture for segmentation of kidney and breast histopathology images. *Comput Electr Eng* 92:107177

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.