



Efficient and robust deep learning architecture for segmentation of kidney and breast histopathology images[☆]

Amit Kumar Chanchal^a, Aman Kumar^a, Shyam Lal^{a,*}, Jyoti Kini^{b,*}

^a Department of Electronics & Communication Engineering, National Institute of Technology Karnataka, Surathkal, Mangalore 575025, Karnataka, India

^b Department of Pathology, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Manipal, India



ARTICLE INFO

Keywords:

Deep learning
Pyramid pooling
Separable convolution
Kidney cancer
Breast cancer
Nuclei segmentation

ABSTRACT

Image segmentation is consistently an important task for computer vision and the analysis of medical images. The analysis and diagnosis of histopathology images by using efficient algorithms that separate hematoxylin and eosin-stained nuclei was the purpose of our proposed method. In this paper, we propose a deep learning model that automatically segments the complex nuclei present in histology images by implementing an effective encoder-decoder architecture with a separable convolution pyramid pooling network (SCPP-Net). The SCPP unit focuses on two aspects: first, it increases the receptive field by varying four different dilation rates, keeping the kernel size fixed, and second, it reduces the trainable parameter by using depth-wise separable convolution. Our deep learning model experimented with three publicly available histopathology image datasets. The proposed SCPP-Net provides better experimental segmentation results compared to other existing deep learning models and is evaluated in terms of F1-score and aggregated Jaccard index.

1. Introduction

With the advancement of image processing and computer vision technology, the manual diagnosis of histopathology images has replaced automated analysis due to the inability to assess large amounts of nuclei present in histopathology images. The automatic segmentation technique provides an authentic and accurate solution to variations in the appearance, texture, and heterogeneous structure of such image slides. Hematoxylin stains cell cores as blue, while eosin stains cytoplasm and other connected tissues as pink. Our segmentation step mainly focused on the separation of nuclei from the cytoplasm. As per [1], it has been discovered that the state of cancer cells may be bigger or smaller than ordinary cells, and that the nuclei of cancerous cells are larger than normal cells and deviate from the center of mass. The nucleus of a cancer cell is darker than that of a normal cell. A pathologist report [2] affirmed that the grade of a cancer can be determined by the size, shape, and other morphology of the associated tissue. The stage or extent of cancer in the body is determined by tumor size, location, and spread. For our purposes, we collected three publicly available datasets of kidney, triple negative breast cancer dataset (TNBC), and multiple organ multi-disease histopathology datasets. Each whole slide image consisted of more than tens of thousands of nuclei with inherent texture differences and distinct biological and clinical behavior. The relevance and potential of the automatic framework for the segmentation of histopathology images has been demonstrated by many authors, but the reported result is still sub-optimal for clinical use. Recent algorithms have

[☆] This paper is for regular issues of CAEE. Reviews processed and approved for publication by the co-Editor-in-Chief Huimin Lu.

* Corresponding authors.

E-mail addresses: shyam.mtec@gmail.com (S. Lal), kini.jyoti@gmail.com (J. Kini).

made numerous new calculations for fusing several deep features to obtain optimal accuracy. The number of parameters and the training time needed to be optimized.

By taking into account these challenges in the segmentation of histopathology images, our study contributes as follows:

(i) To extract the high-level feature from complex structured histopathology images, our SCPP network leverages the strength of the encoder-decoder as well as pyramid pooling at the bottleneck layer.

(ii) For the fast assessment, we used depth-wise separable convolution that effectively reduced the total number of parameters and floating point operations per second (FLOPs) in the SCPP-Net.

The organization of the paper is as follows. In Section 2, we describe the related research work. Section 3 covers the core of the findings and describes them with the help of mathematical equations and block diagram. Implementation, training, and detailed architecture are presented in Section 4. Visual and tabular representation of results are shown in Section 5, and lastly, Section 6 presents the conclusions of our work.

2. Related work

The deep neural network has achieved a dominant breakthrough in the image processing area and computer vision task. One of the most popular convolutional neural networks for biomedical image segmentation is known as the U-Net, developed by [3]. U-Net architecture consists of down-sampling and up-sampling paths and yields a U-shaped network. The original U-Net does not use batch normalization, skipping connections, and many other potential opportunities for further enhancements in segmentation applications. For the automated segmentation of breast tissue, [4] proposed sparse reproduction, multi-layer convolution systems, and morphological operations. [5]'s CNN model used a bottom-up and top-down shape deformation approach to precisely segment the nuclei, where they combined the region iteratively and merged it for better initialization of tissue morphology. SegNet architecture in [6] is a deep CNN for road and indore segmentation. SegNet network has an encoder path to produce a sparse feature map, a corresponding decoder network followed by a classifier (Softmax), which is used for pixel-wise classification. Attention U-Net in [7] incorporated an attention gate that has the capability of automatic learning, and the focus was to learn the variation in dimension consistently. With FCN and atrous spatial pyramid pooling, [8] re-introduced the input image within the network at several points to minimize the degradation of context in an image. For instant nucleus segmentation, [9] proposed a network that has an encoder, decoder, and a series of residual blocks. To improve the segmentation process, [10] used a combined loss function and transfer learning. [11] solved the segmentation problem as a regression task by calculating the distance between nuclei and formulating a global loss function that precisely segments the touching nuclei. [12] achieved significant improvements compared to other existing models by utilizing a new convolutional unit called dimension-wise convolution to learn the spatial representation of channels in a better manner. For the successful separation of overlapped nuclei, [13] used the synchronous segmentation method to calculate the vertical and horizontal separations of nuclear pixels to their centers of mass. For automatic segmentation of ductal carcinoma in situ in whole slide images, [14] used a standard segmentation model UNet with ELU activation and batch normalization. In [15], a multi-scale structure precisely segments the glands in histological images where the network has a fully convolutional residual structure with a dilated convolution. For gland and lumen segmentation in colon histology, [16] introduced a rotation equivalent network that uses group-equivalent convolutional neural networks (G-CNN) having an additionally rotated kernel. To capture the multilayer features of complex nuclei, [17]'s encoder-decoder network used atrous spatial pyramid pooling, a concave point detection method, a combination of dice loss, and binary cross-entropy loss as a joint loss function. By assimilating residual block, attention block, and bottleneck layer, [18] incorporated three efficient blocks in encoder-decoder architecture to regain the meaningful information lost during max-pooling operation and correct localization of objects in the decoder section.

3. Proposed SCPP-Net architecture

Commonly used architecture for image segmentation is based on encoder and decoder architectures, such as [3,7]. The working method of our proposed SCPP-Net model was based on the principle of extracting more relevant features at a higher level. Our SCPP layer increased the receptive field by varying four different dilation rates and keeping a fixed kernel size. Dilation rate was an additional parameter that varied to the resultant feature maps to visualize larger areas. This is advantageous in the sense that we added little computational complexity and were able to obtain a larger receptive field. One of the significant challenges in the segmentation of nuclei present in the histopathology image is the separation of clumped and overlapped nuclei. Due to this closeness and overlapping of nuclei, it is difficult to recognize them separately. We performed the segmentation using the proposed SCPP-Net model, which is CNN-based architecture and inspired by [7,17]. This CNN-based architecture is helpful for overcoming the problem of closeness and overlapping nuclei by increasing the receptive field at a higher level.

3.1. Description of SCPP-Net architecture

Based on [3], we conceived of the idea of an encoder and decoder. In the down-sampling path, we performed convolution and max-pooling operations for the input images. In this path, the focus was on capturing the context in the image. In the contraction path, the image size gradually reduced, while the depth of the image gradually increased. In the deep layer of the convolution operation, our SCPP block was used to extract more relevant features to increase the receptive field by varying four different dilation rates and keeping a fixed kernel size. In the expanding path, the size of the image gradually increased and the depth gradually diminished. Decoder path enabled precise localization by gradually applying up-sampling. For each step in the expanding path, we

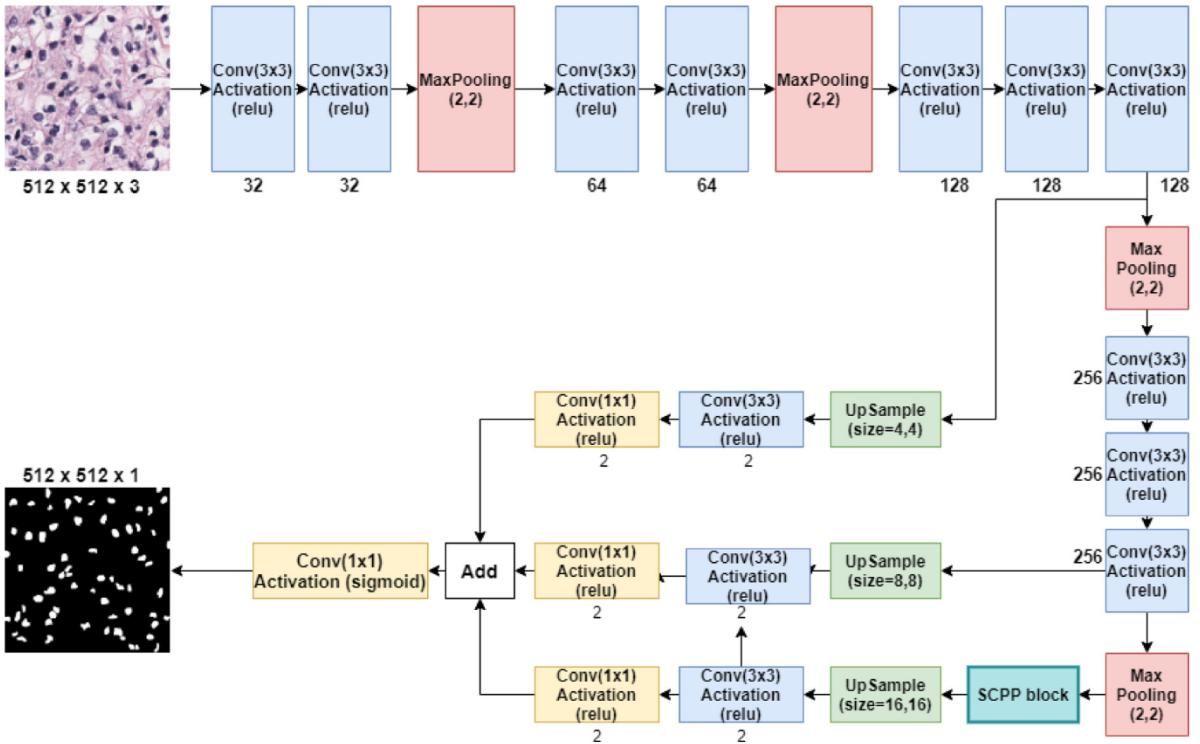


Fig. 1. Detailed diagram of proposed Separable convolution pyramid pooling network (SCPP-Net).

utilized a skip connection by concatenating the output of the decoder convolution layers with feature maps from the contraction path at a similar level. This was an end-to-end completely convolutional system, and it did not contain any dense layer. This was invariant to image size, which means it can accept any size of input image. The detailed diagram of the proposed SCPP-Net model is presented in Fig. 1, and a layer-wise summary of the proposed SCPP-Net model is presented in Table 1.

3.2. Atrous spatial pyramid pooling

In atrous spatial pyramid pooling (ASPP), atrous convolution is used with different dilation rates for a better receptive field of view. Atrous convolution is also known as dilation convolution. As described in [19], assuming a one-dimensional signal, the output $y[n]$ of the atrous convolution of a one-dimensional input signal $x[n]$ with a filter $w[n]$ of length M is defined as:

$$y[n] = \sum_{i=1}^M x[n + r.i]w[i] \quad (1)$$

For each spatial position n on the output y and kernel w , atrous convolution is applied to input x . Convolution of input x with up-sampled filters produced by inserting $(r - 1)$ zero between two consecutive filters produced one-dimensional output of length M . Each location of (m, n) of the output y atrous convolutions applied to input x results in a two-dimensional image of size $M \times N$ given in Eq. (2). In Eqs. (1) and (2), r corresponds to the dilation rate with which the input signal goes into the sample. When $r = 1$, a normal convolution operation occurs and we used normally CNN networks. When $r > 1$, then it allows us to expand our filter strides with different rates, so this convolution becomes atrous or dilation convolution, which increases receptive field of view.

$$y(m, n) = \sum_{i=1}^M \sum_{j=1}^N x(m + r.i, n + r.j)w(i, j) \quad (2)$$

The receptive field (global view) visualization of atrous convolution with dilated rates, $r = 2, 4, 6$, and 8 , is shown in Fig. 2. The selection of dilation rates was based on obtaining optimum accuracy while keeping the number of parameters minimum. In Fig. 2, the green color indicates how the receptive field increases with different dilation rates. ASPP is a version of spatial pyramid pooling, which is used in [20]. We convolved the input feature map parallelly with the different dilation rates (r) and fused them. ASPP helped us identify the object of the same class with different scales in the image. So, with the identification of the different object scales, our accuracy was also improved. With this motivation, we used ASPP, but instead of a normal convolution, we used separable convolution, as used in [21]. Application of L filters of size $S_k \times S_k \times 1$ on feature size $S_f \times S_f \times L$ yields feature size of $S_p \times S_p \times L$. Convolution of N filters of $1 \times 1 \times L$ with $S_p \times S_p \times L$ results $S_p \times S_p \times N$. By this way total number of multiplications in depth-wise

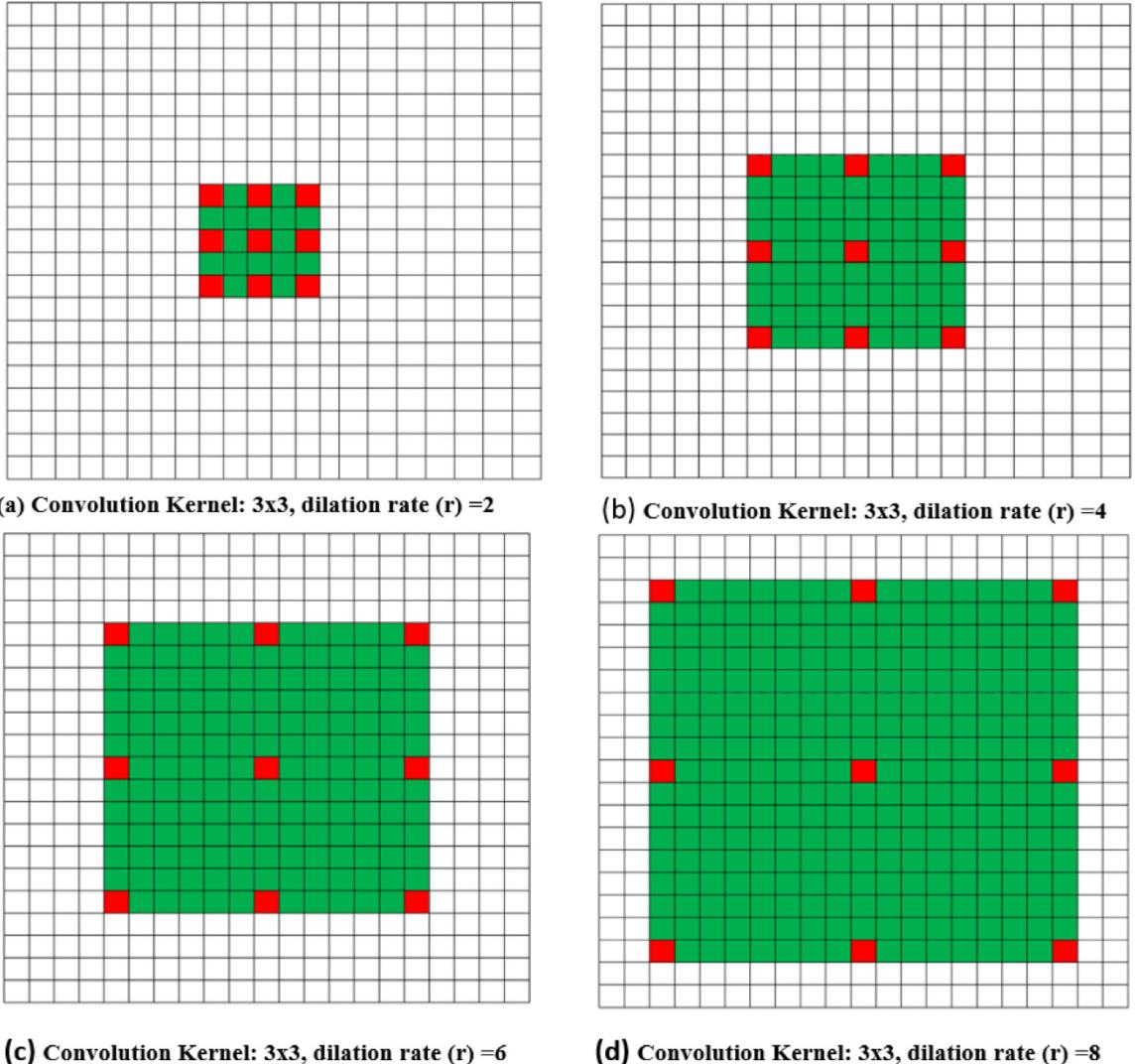


Fig. 2. Visualization of the receptive field (global view) of atrous convolution with different dilation rates.

separable convolution is $(S_k \times S_k \times 1)(S_p \times S_p \times L) + (1 \times 1 \times L)(S_p \times S_p \times N)$, which is significantly less than standard convolution. Visualization of detailed computational complexity in standard convolution and depthwise separable convolution is shown in Fig. 3. ASPP used separable convolution with 3×3 kernel sizes and dilation rates of 2, 4, 6, 8. We concatenated all of the above with the same depth and convolved (separable) with 1×1 kernel size so that the model was able to extract maximum features. The decoder path was pipelined in such a way that it up-sampled all the feature maps parallelly. Thus, we obtained the same size of the image. The block diagram shown in Fig. 4 describes the detailed architecture of the proposed SCPP-Net.

4. Training and implementation

We have conducted our training and implementation in Google Colab GPU with the latest version of Keras and TensorFlow Python-3 framework to speed up the development process. Before performing the training, we reviewed our dataset to test its validity with the proper data and labels. In the process of training, we performed some data augmentation as well as pre-processing (for the TNBC dataset and multiple organs multi-disease). We trained our model on the training set for a given range of hyperparameters and evaluated the efficiency of the model on the test and validation datasets. A summary of our CNN-based architecture is shown in Table 1. For a given range of hyperparameters, we repeated this process until we obtained the best result in terms of the F1 score and Jaccard index. The source code of proposed SCPP-Net model is available at <https://github.com/shyamfec/SCPP-Net>

We used Adam described in [22] as an optimizer, and our batch size was four (for kidney dataset) and one (for breast dataset and multiple organs multi-disease dataset) for the best possible result. These are the hyperparameters that we varied: (1) learning rate α , (2) weight decay constant λ , (3) the size of the convolution filter n_f , (4) dilation rate, and (5) batch size

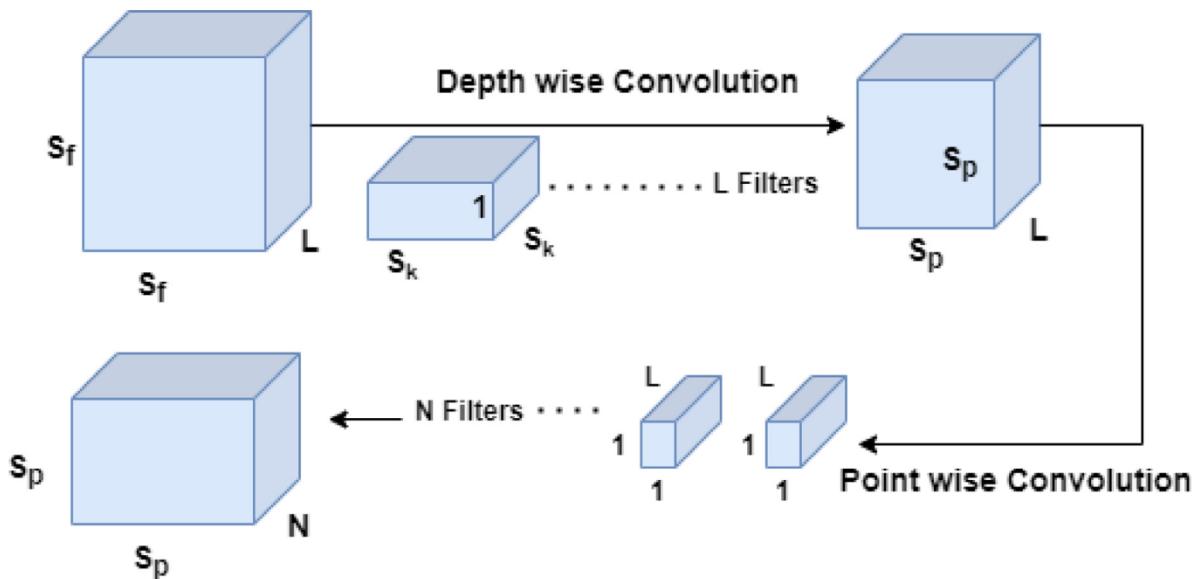


Fig. 3. Visualization of depth-wise separable convolution.

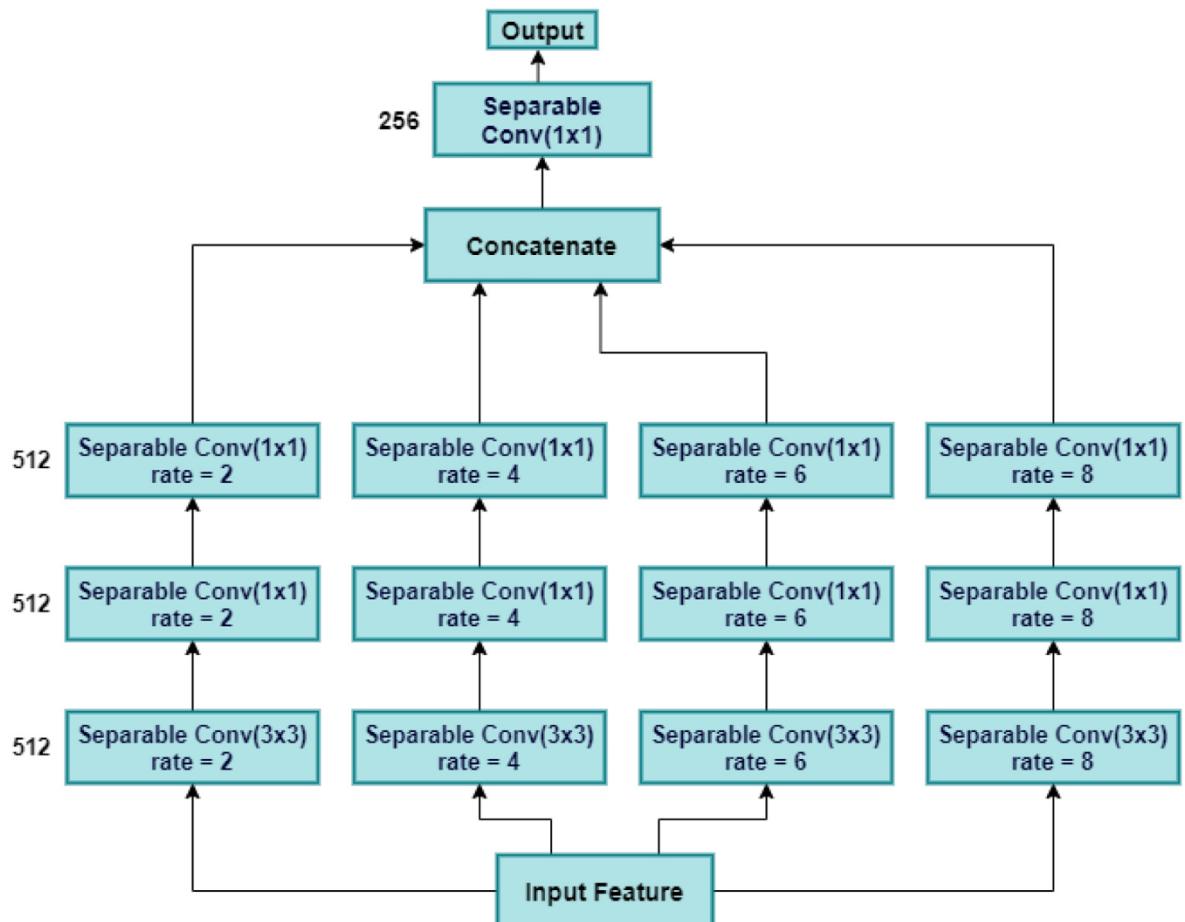


Fig. 4. Separable convolution pyramid pooling block.

Table 1
Layerwise summary of proposed SCPP-Net architecture.

Layers	Size_In	Size_Out	Kernel_size
Conv1	512 × 512 × 3	512 × 512 × 32	3 × 3 × 3 × 32
Conv1	512 × 512 × 32	512 × 512 × 32	3 × 3 × 32 × 32
Pool1	512 × 512 × 32	256 × 256 × 32	2 × 2 × 32 × 32
Conv2	256 × 256 × 32	256 × 256 × 64	3 × 3 × 32 × 64
Conv2	256 × 256 × 64	256 × 256 × 64	3 × 3 × 64 × 64
Pool2	256 × 256 × 64	128 × 128 × 64	2 × 2 × 64 × 64
Conv3	128 × 128 × 64	128 × 128 × 128	3 × 3 × 64 × 128
Conv3	128 × 128 × 128	128 × 128 × 128	3 × 3 × 128 × 128
Conv3	128 × 128 × 128	128 × 128 × 128	3 × 3 × 128 × 128
Pool3	128 × 128 × 128	64 × 64 × 128	2 × 2 × 128 × 128
Conv4	64 × 64 × 128	64 × 64 × 256	3 × 3 × 128 × 256
Conv4	64 × 64 × 256	64 × 64 × 256	3 × 3 × 256 × 256
Conv4	64 × 64 × 256	64 × 64 × 256	3 × 3 × 256 × 256
pool4	64 × 64 × 256	32 × 32 × 256	2 × 2 × 256 × 256
ASPP	32 × 32 × 256	32 × 32 × 256	
UpSample1	32 × 32 × 256(conv5)	512 × 512 × 256	16 × 16 × 256 × 256
UpSample2	64 × 64 × 256(conv4)	512 × 512 × 256	8 × 8 × 256 × 256
UpSample3	128 × 128 × 128(conv3)	512 × 512 × 128	4 × 4 × 128 × 128
Conv6	512 × 512 × 256(upsample1)	512 × 512 × 2	3 × 3 × 256 × 2
Conv7	512 × 512 × 256(upsample2)	512 × 512 × 2	3 × 3 × 256 × 2
Conv8	512 × 512 × 256(upsample3)	512 × 512 × 2	3 × 3 × 256 × 2
Conv9	512 × 512 × 2(conv6)	512 × 512 × 2	1 × 1 × 2 × 2
Conv10	512 × 512 × 2(conv7)	512 × 512 × 2	1 × 1 × 2 × 2
Conv11	512 × 512 × 2(conv8)	512 × 512 × 2	1 × 1 × 2 × 2
	512 × 512 × 2(conv9)		
Add	512 × 512 × 2(conv10)	512 × 512 × 2	
	512 × 512 × 2(conv11)		
Conv12	512 × 512 × 2	512 × 512 × 1	1 × 1 × 2 × 1

According to [23], batch size and learning rates have a significant impact on the model performance. We had 654 training images in the kidney dataset without any data augmentation. After data augmentation, the TNBC and multi-organ datasets had 200 and 288 training images, respectively. We chose a learning rate of 0.0001 for training all the models with three datasets. Selection of learning rate and weight decay constant was based on trial and error, as it cannot be calculated analytically. As there was a strong correlation between batch size and learning rate, our model yielded satisfactory results for batch size four for the kidney dataset and one for the other two cases. The selection of those two hyperparameters, such as batch size and learning rates, sped up the training process. Our model also utilized flexible size convolution filters based on available resources.

4.1. Description of histopathology dataset

(I) Hematoxylin and eosin (H&E)-stained triple negative breast cancer (TNBC) dataset: This dataset was obtained from [11]. This dataset was a corpus of 50 H&E-stained histopathology images of human breast tissue and consisted of 4022 cells that were accurately annotated by a pathologist.

(II) H&E stained kidney dataset: This dataset was obtained from [24]. This corpus of kidney renal clear cell carcinoma images was obtained from the National Cancer Institute. We obtained a subset of 730 H&E-stained histopathology images for our task of nuclei segmentation.

(III) Multiple organs multi-disease histopathology dataset: This dataset was prepared by [25] and included seven different organs of kidney, liver, colon, breast, prostate, bladder, and stomach, with a total of 26 747 annotated nuclei.

4.2. Data augmentation and preprocessing

To increase the diversity of data for training, the strategy of data augmentation by the creation of artificial variations, such as rotation, flipping, padding, and zooming, dataset were expanded. The purpose of augmentation is to prevent CNN from undergoing overfitting, and the augmentation approach depends on the type of architecture used. In our case, we used different types of data augmentation for the TNBC and multiple organ multi-disease histopathology datasets, so that we obtained more images for the training set as well as the validation set. We used flipping, rotation, dihedral, and random lighting (the threshold is 0.5). We performed color normalization for the TNBC and multi-disease histopathology datasets. Color normalization is needed to match the color characteristics of an image to those of the target image. In this process, we matched the mean and standard deviation of each color channel in the two images.

Table 2
Performance measure of proposed model with TNBC dataset.

Model	F1 score	AJI score	Parameters	FLOPs
Proposed SCPP-Net (without separable)	0.7661	0.6551	9,267,899	18,520,320
Proposed SCPP-Net (with separable)	0.8168	0.6998	5,088,955	10,162,445

Table 3
Performance measure of proposed model with kidney dataset.

Model	F1 score	AJI score	Parameters	FLOPs
Proposed SCPP-Net (without separable)	0.8975	0.8290	9,267,899	18,520,320
Proposed SCPP-Net (with separable)	0.9203	0.8581	5,088,955	10,162,445

Table 4
Performance comparison of architectures with kidney dataset.

Model	F1 score	AJI score	Parameters	FLOPs
U Net(2015)	0.8537	0.7489	31,378,945	62,746,139
Segnet(2017)	0.8972	0.8304	18,826,753	47,065,663
Attention UNet(2019)	0.9135	0.8590	31,902,629	63,791,573
DIST(2019)	0.8992	0.8272	7,771,873	15,525,181
ASPPU-Net(2020)	0.9052	0.8293	4,307,713	8,610,587
Proposed SCPP-Net	0.9203	0.8592	5,088,955	10,162,445

4.3. Loss function

Learning is driven by errors; to guide the learning algorithm, the aim is to minimize the loss, which is done by an effective loss function. For the regression problem, the squared error, the loss is plausible because we predicted some value, and we were curious to see how far we from those values. For classification, we realized that it was a distribution, so maybe we could use something that allows us to capture the difference between the true distribution and the predicted distribution. In this work, our loss function is the binary cross-entropy (BCE) used in [3,18].

$$\text{Cross-Entropy Loss} = - \sum_{k=1}^C y_k \log(f(s)_k) \quad (3)$$

where y_k and s_k are ground truth and predicted score for each class k in C . Eq. (3) is cross-entropy loss for C classes. Typically, we use relu activation in the middle layer and sigmoid activation before the cross entropy for loss computation. $f(s_k)$ indicates the activation unit and in most of the literature it is called \hat{y} . Eq. (4) represents binary cross-entropy loss, since it is similar to a binary classification problem for $C' = 2$ classes. Cross-entropy loss followed by sigmoid activation is called sigmoid cross-entropy loss, and similarly, cross-entropy loss followed by softmax activation is called softmax cross-entropy loss. We had two different symbols, C and C' , in two different equations because in the first case, C indicates more than two classes and multi-labels. Eq. (5) indicates the binary classification for the two labels.

$$\text{BCE Loss} = - \sum_{k=1}^{C'=2} y_k \log(f(s)_k) \quad (4)$$

$$\text{BCE Loss} = -[y \ln(\hat{y}) + (1 - y) \ln(1 - \hat{y})] \quad (5)$$

4.4. Ablation study

To demonstrate the effectiveness of the proposed SCPP-Net architecture, an ablation study was performed on two publicly available kidney and TNBC histopathology datasets. In this study, the effectiveness of each module was demonstrated quantitatively and qualitatively. The results of the proposed SCPP-Net architecture, with and without separable convolution, are reported in Tables 2 and 3 and in Fig. 5.

5. Experimental segmentation results

After the simulation, we were able to compute and analyze our experimental results of nuclei segmentation on three publicly available H&E-stained histopathology image datasets. We used the image quality metrics F1 score used in [11,18] and aggregated Jaccard index (AJI) scores used in [11] to quantify the effectiveness of the proposed SCPP-Net architecture and other existing deep learning segmentation methods. The number of parameters and FLOPs quantified the GPU computing power of the proposed SCPP-Net architecture and other existing deep-learning segmentation methods.

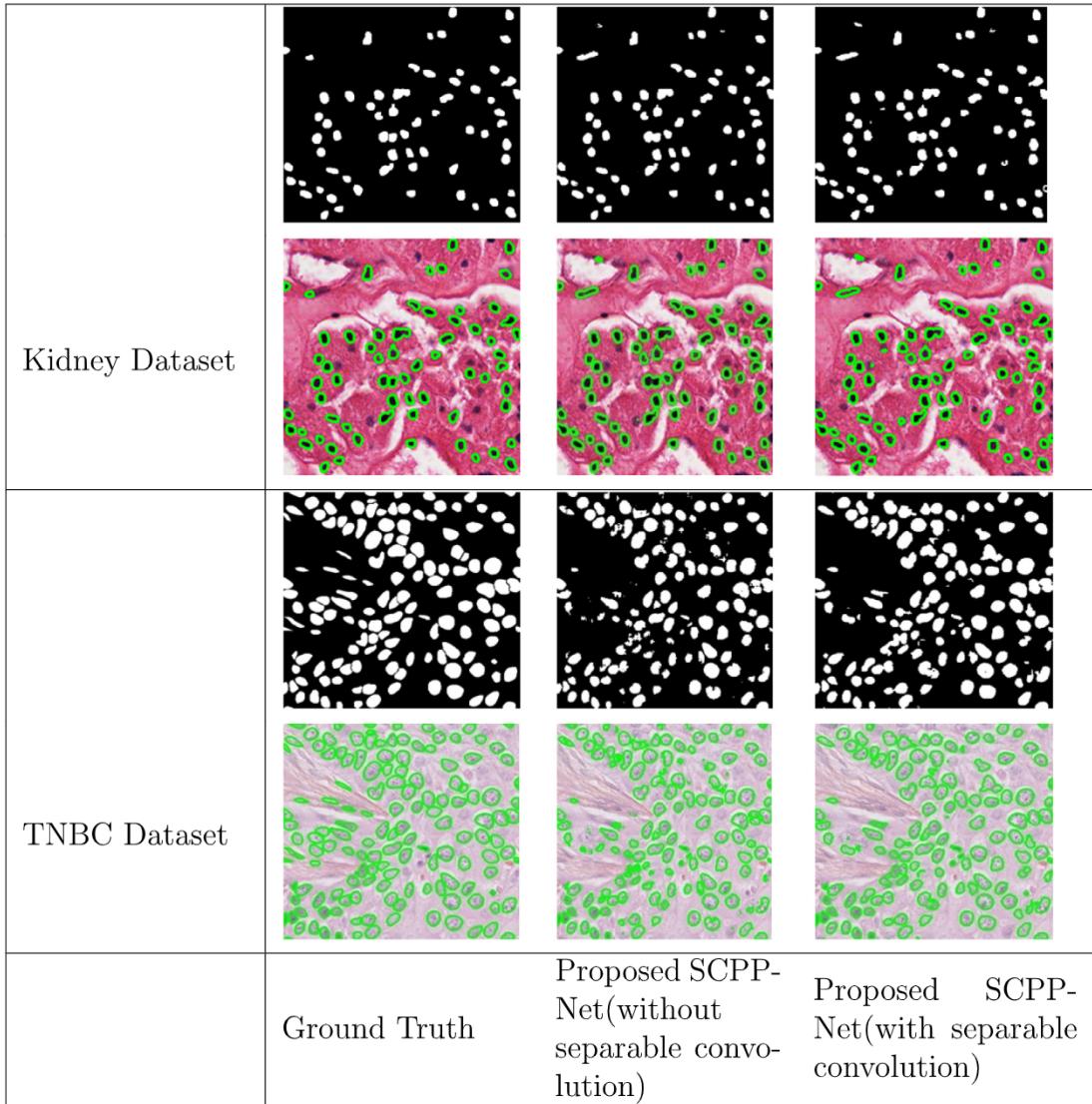


Fig. 5. Binary and overlay images of proposed SCPP-Net architecture (with and without separable convolution) on kidney and TNBC datasets.

Table 5
Performance comparison of architectures with TNBC dataset.

Model	F1 score	AJI score	Parameters	FLOPs
U-Net(2015)	0.7324	0.6559	31,378,945	62,746,139
Segnet(2017)	0.7685	0.6434	18,826,753	47,065,663
Attention UNet (2019)	0.7216	0.6194	31,902,629	63,791,573
DIST(2019)	0.7516	0.6727	7,771,873	15,525,181
ASPPU-Net(2020)	0.7781	0.6378	4,307,713	8,610,587
Proposed SCPP-Net	0.8168	0.6998	5,088,955	10,162,445

5.1. Comparison of state-of-the-art segmentation methods

To demonstrate the effectiveness of the proposed SCPP-Net architecture with other existing deep learning segmentation methods [3,6,7,11,17], nuclei segmentation results were evaluated and compared on three publicly available kidney, TNBC, and multiple organs multi-disease histopathology image datasets. **Table 4** shows a quantitative segmentation results comparison of the proposed SCPP-Net architecture with five other existing deep-learning segmentation architectures on the kidney histopathology dataset. Performance measurement in terms of F1 Score, AJI, the total number of trainable parameters, and FLOPs described the

Table 6
Performance comparison of architectures with multiple organs multi-disease dataset.

Model	Avg. F1 score of 7 different organs	Avg. AJI score of 7 different organs	Parameters	FLOPs
U-Net(2015)	0.7681	0.6295	31,378,945	62,746,139
Segnet(2017)	0.7846	0.6488	18,826,753	47,065,663
Attention UNet (2019)	0.7538	0.6098	31,902,629	63,791,573
DIST(2019)	0.7831	0.6480	7,771,873	15,525,181
ASPPU-Net(2020)	0.7839	0.6483	4,307,713	8,610,587
Proposed SCPP-Net	0.8010	0.6710	5,088,955	10,162,445

training time and computational complexity. **Table 5** shows a quantitative segmentation results comparison of the proposed SCPP-Net architecture with five other existing deep-learning segmentation architectures on the TNBC histopathology dataset. **Table 6** shows a quantitative segmentation results comparison of the proposed SCPP-Net architecture with five other existing deep-learning segmentation architectures on the multiple organ multi-disease histopathology dataset. The results indicated that our proposed SCPP-Net architecture was able to retrieve more information compared to other existing deep learning segmentation architectures.

5.2. Computational complexity analysis

The computational complexity of the proposed SCPP-Net architecture and existing deep learning segmentation methods (i.e., U-Net [3], SegNet [6], Attention UNet [7], Dist [11], and ASPPU-Net [17]) were analyzed in terms of the number of trainable parameters and FLOPs. The value of FLOPs describes the computing power of given hardware, such as the GPU; the smaller the value, the faster the computing ability. The trainable parameters and FLOPs of the proposed SCPP-Net architecture and existing deep learning segmentation methods are presented in the fourth and fifth columns of **Tables 4**, **5**, and **6**. From the tables, it is clear that the proposed SCPP-Net required fewer trainable parameters and FLOPs compared to the most recent deep learning segmentation methods because we used ASPP with separable convolution. Training time was proportional to the number of trainable parameters and the size of the training dataset. In this process, the average training time of the proposed model was 27 s, 17 s, and 14 s per epoch for the kidney, TNBC, and multi-organ datasets, respectively. We terminated the training process when validation accuracy was not improved in 10 successive iterations. Based on the improvement in validation accuracy, we were able to train the model an average of 45 to 65 epochs.

5.3. Visual segmentation results

Visual nuclei segmentation results comparison of proposed and different existing deep learning segmentation architectures on kidney dataset, TNBC dataset, and multiple organs multi-disease dataset are shown with two sample images and their overlay images in **Figs. 6**, **7**, and **8** respectively. Number of clustered predictions was less in the case of SegNet as compared to U-Net and Attention U-Net. False detection was very low in the case of SegNet. The number of partially detected nuclei was maximum in the case of Attention U-Net compared to any other model. Partially detected nuclei in Dist were fewer compared to SegNet, U-Net, and Attention U-Net. Prediction accuracy in terms of partially detected nuclei and clearly identified nuclei slightly improved in ASPPU-Net compared to U-Net and SegNet. The problem of separating overlapped nuclei was better handled by Dist. In the proposed model, there were very few false-detected nuclei and partially detected nuclei. The problem of separating clustered nuclei was not completely solved. Visual results indicated that the proposed SCPP model yielded very good results in terms of the number of clearly identified nuclei, the number of absent nuclei, partially detected nuclei, the number of overlapped nuclei, and false detection. A graphical illustration of the computational complexity comparison of nuclei segmentation results of the different CNN models (Model-1 (UNet (2015)), Model-2 (SegNet (2017)), Model-3 (Attention U-Net (2019)), Model-4 (Dist (2019)), Model-5 (ASPPU-Net(2020)) and the proposed SCPP-Net model) is shown in **Fig. 9**.

6. Conclusion

For automatic diagnosis and prognosis of cancer disease from H&E-stained histology data, the design of an efficient and robust computer-aided system is in high demand. In this paper, we presented a new deep learning framework for the nuclei segmentation of histopathology images that addresses two major problems. To separate nuclei from complex structured histopathology images with different histology and molecular characteristics was the first major concern. This issue was addressed by implementing an effective encoder-decoder architecture with the SCPP unit at the bottleneck layer. SCPP block extracted and learned deep features by varying four different dilation rates. The second major concern was to reduce the computational complexity and total trainable parameters, which was effectively handled by incorporating depth-wise separable convolution. To prove the worthiness of our robust network, we carefully experimented with three different histopathology datasets. Quantitative and qualitative results yielded for the three histopathology datasets showed that our SCPP-Net architecture was best in terms of F1 score and AJI score, with less computational complexity. The tabular and visual segmentation results of the proposed framework were better in comparison to other benchmark deep learning models. The measured values of (F1-score, AJI score) were (0.9203, 0.8592), (0.8168, 0.6998), and (0.8010, 0.6710) for kidney, TNBC, and multiple organ multi-disease histopathology datasets, respectively.

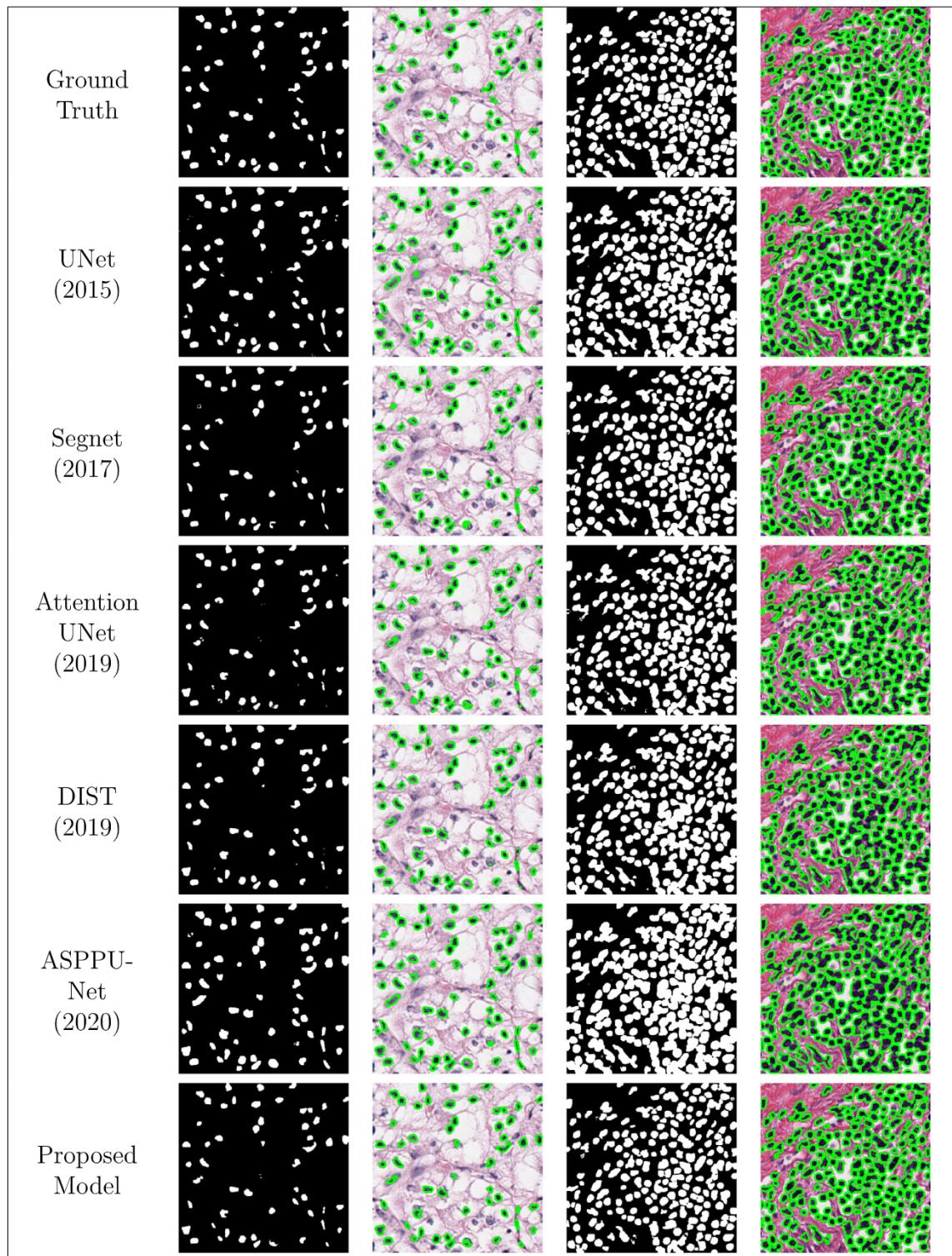


Fig. 6. Visual segmentation results comparison of different models on kidney dataset (column 1- different comparison models, column 2 & 4-ground truth & predicted images, column 3 & 5-corresponding overlay images).

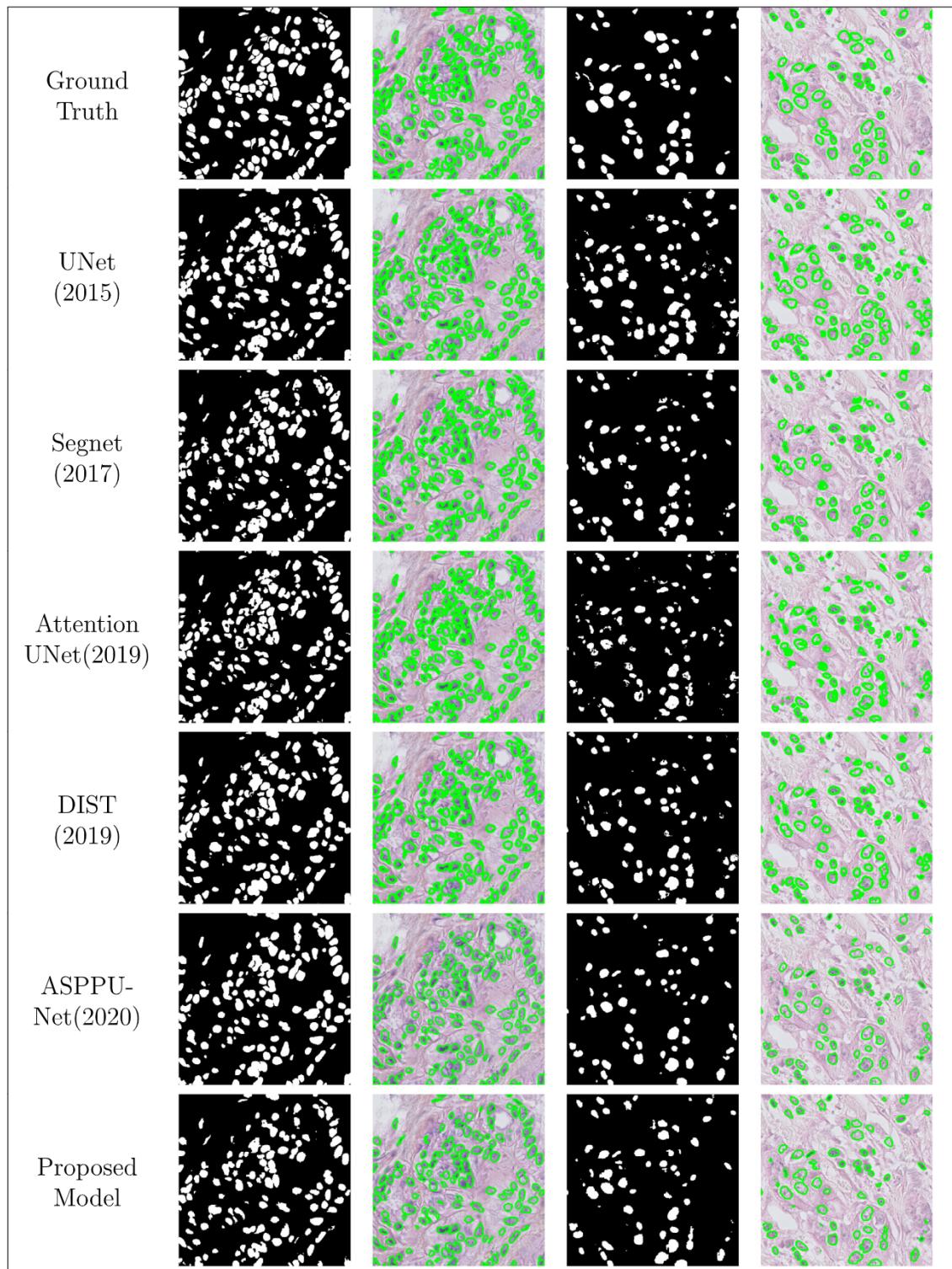


Fig. 7. Visual segmentation results comparison of different models on TNBC dataset (column 1- different comparison models, column 2 & 4-ground truth & predicted images, column 3 & 5-corresponding overlay images).

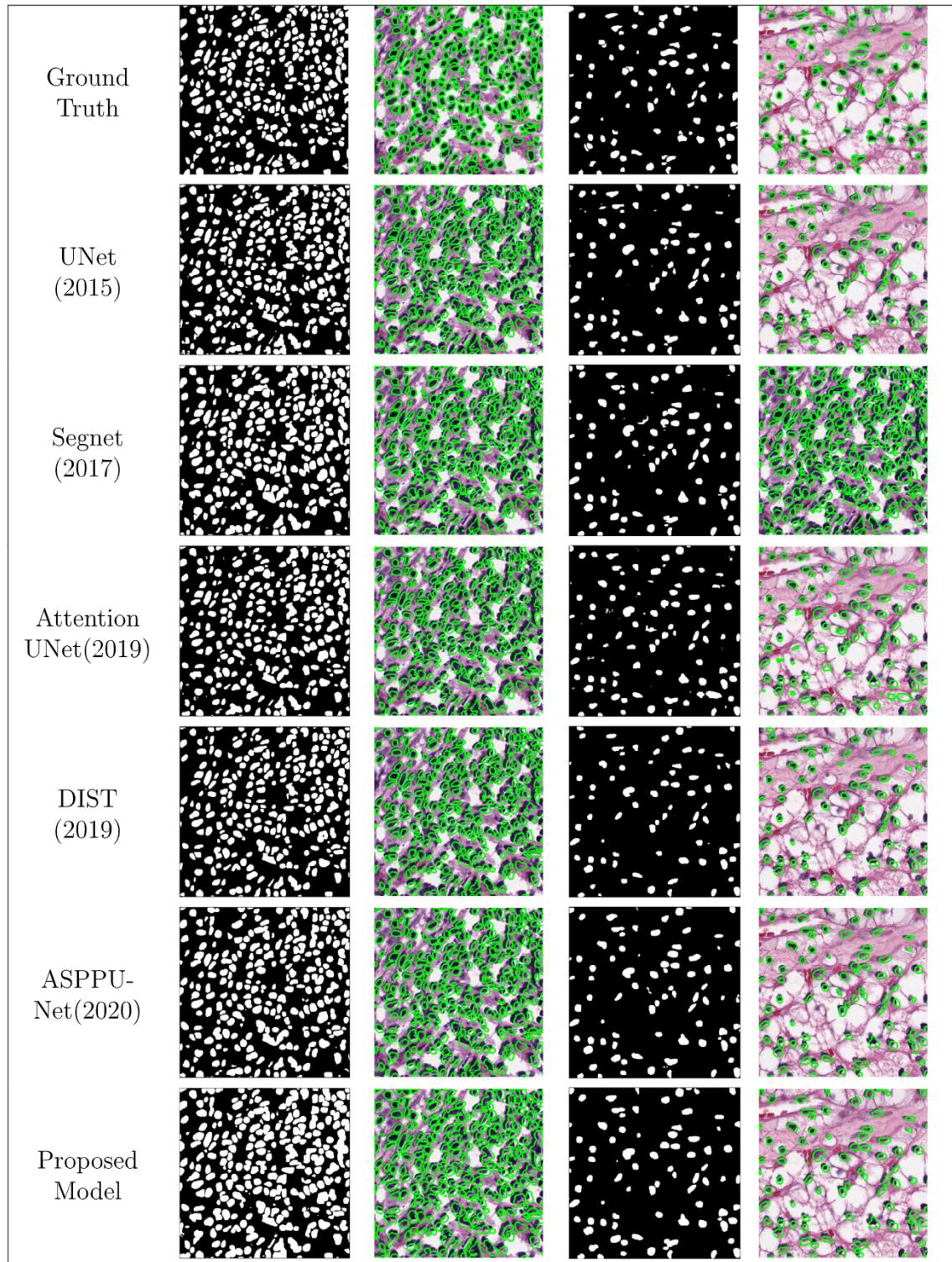


Fig. 8. Visual segmentation results comparison of different models on multiple organ multi-disease dataset (column 1- different comparison models, column 2 & 4-ground truth & predicted images, column 3 & 5-corresponding overlay images).

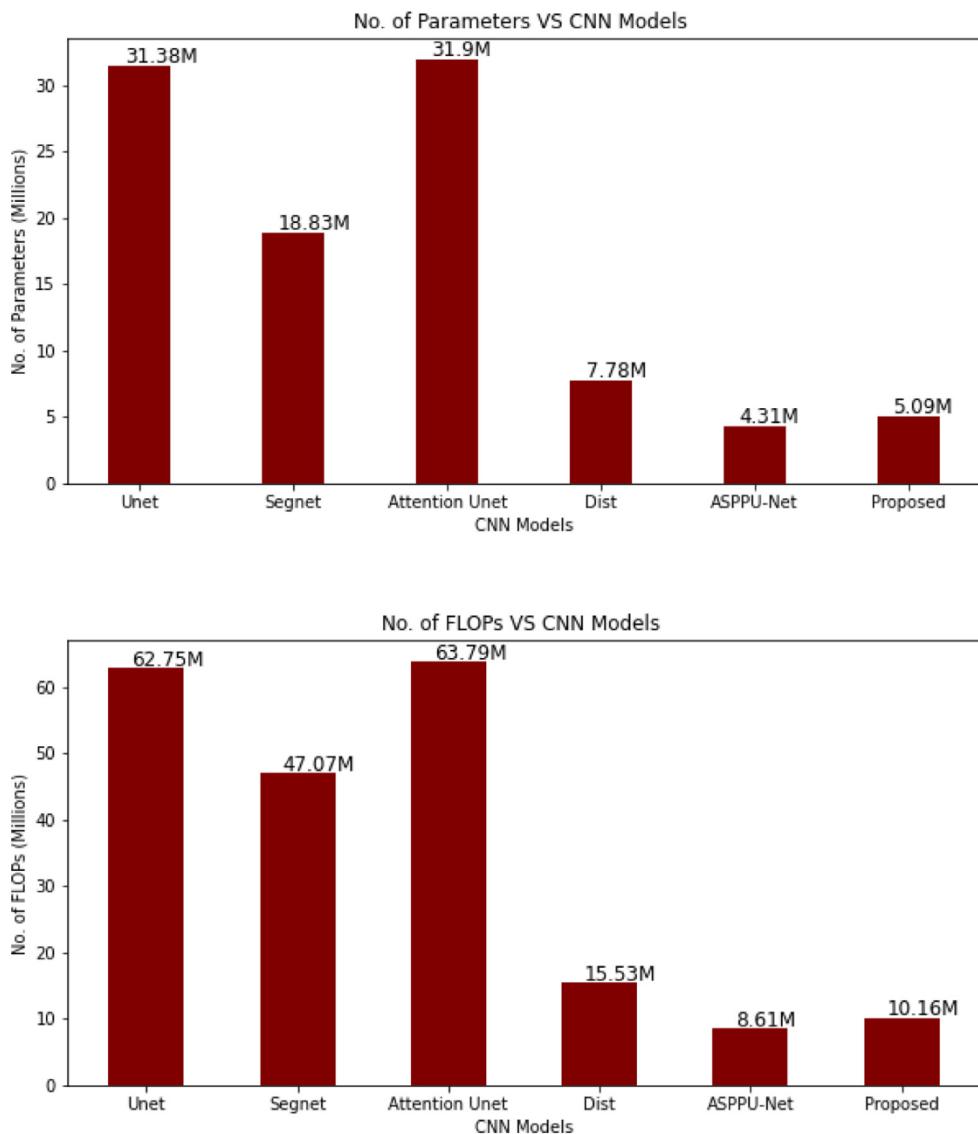


Fig. 9. Computational complexity comparison of different CNN models.

All the work in this study focused on the separation of hematoxylin-stained cell nuclei from eosin cells. Pathologist reports have indicated that in the case of higher-grade cancer, eosin of the cytoplasm regions also need to be analyzed. Another limitation is the segmentation of overlapped nuclei, which has not been solved completely. These issues will be the focus of our future work.

CRediT authorship contribution statement

Amit Kumar Chanchal: Conception and design of study, Writing - original draft, Writing - review & editing. **Aman Kumar:** Conception and design of study, Writing - original draft, Writing - review & editing. **Shyam Lal:** Acquisition of data, Analysis and/or interpretation of data, Writing - original draft, Writing - review & editing. **Jyoti Kini:** Acquisition of data, Analysis and/or interpretation of data, Writing - review & editing.

Declaration of competing interest

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to <https://doi.org/10.1016/j.compeleceng.2021.107177>.

Acknowledgments

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. All authors approved the version of the manuscript to be published.

References

- [1] Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: Experience from a large study with long-term follow-up. *Histopathology* 1991;19:403–10.
- [2] 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. Protocol for the examination of specimens from patients with primary carcinoma of the colon and rectum. *Arch Pathol Lab Med* 2009;133:1539–51.
- [3] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: Proc. MICCAI. Munich, Germany: Springer; 2015, p. 234–41.
- [4] Pan X, Li L, Yang H, Liu Z, Yang J, Zhao L, Fan Y. Accurate segmentation of nuclei in pathological images via sparse reconstruction and deep convolutional networks. *Neurocomputing* 2017;229:88–99.
- [5] Xing F, Xie Y, Yang L. An automatic learning-based framework for robust nucleus segmentation. *IEEE Trans Med Imaging* 2016;35(2):550–66.
- [6] Badrinarayanan V, Kendall A, Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell* 2017;39(12):2481–95.
- [7] Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B, Glocker B, Rueckert D. Attention U-Net: Learning where to look for the pancreas. In: 1st conference on medical imaging with deep learning MIDL 2018. Amsterdam Netherlands, <https://arxiv.org/abs/1804.03999>.
- [8] Graham S, Chen H, Dou Q, Heng PA, Snead D, Tsang YW, Rajpoot N. MILD-Net: Minimal information loss dilated network for gland instance segmentation in colon histology images. *Med Image Anal* 2018;52:199–211.
- [9] Graham S, Rajpoot NM. SAMS-NET: Stain-aware multi-scale network for instance-based nuclei segmentation in histology images. In: 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018). Washington, DC. 2018, p. 590–4.
- [10] Qu H, Riedlinger G, Wu P, Huang Q, Yi J, De S, Metaxas D. Joint segmentation and fine-grained classification of nuclei in histopathology images. In: 2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019). Venice, Italy. 2019, p. 900–4.
- [11] Naylor P, Lae M, Reyal F, Walter T. Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE Trans Med Imaging* 2019;38(2):448–59.
- [12] Mehta S, Hajishirzi H, Rastegari M. DiCENet: Dimension-wise convolutions for efficient networks. In: Computer vision and pattern recognition, Vol. 3. 2020, p. 1–12, <https://arxiv.org/abs/1906.03516>.
- [13] Graham S, Dang Q, Raza SA, Azam A, Tsang YW, Kwak JT, Rajpoot N. HoVer-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Med Image Anal* 2019;58:101563, 1–15.
- [14] Seth N, Akbar S, Nofech MS, Salama S, Martel A. Automated segmentation of DCIS in whole slide images. In: Digital pathology ECDP 2019. Lecture notes in computer science, vol. 11435, Springer, Cham.
- [15] Ding H, Pan Z, Cen Q, Li Y, Chen S. Multi-scale fully convolutional network for gland segmentation using three class classification. *Neurocomputing* 2019;380:150–61.
- [16] Graham S, Epstein D, Rajpoot N. Rota-Net: Rotation equivariant network for simultaneous gland and lumen segmentation in colon histology images. In: European congress on digital pathology. Cham: Springer; 2019, p. 109–16.
- [17] Wan T, Zhao L, Feng H, Li D, Tong C, Qin Z. Robust nuclei segmentation in histopathology using ASPPU-Net and boundary refinement. *Neurocomputing* 2020;408:144–56. <http://dx.doi.org/10.1016/j.neucom.2019.08.103>.
- [18] Lal S, Das D, Alahya K, Kanfade A, Kumar A, Kini J. NucleiSegNet: Robust deep learning architecture for the nuclei segmentation of liver cancer histopathology images. *Comput Biol Med* 2021;128:104075. <http://dx.doi.org/10.1016/j.combiomed.2020.104075>.
- [19] Chen LC, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation. In: Computer vision and pattern recognition, Vol. 3. 2017, p. 1–14, <https://arXiv:1706.05587v3> [cs.CV].
- [20] He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans Pattern Anal Mach Intell* 2015;37(9):1904–16. <http://dx.doi.org/10.1109/TPAMI.2015.2389824>.
- [21] He Y, Qian J, Wang J. Depth-wise decomposition for accelerating separable convolutions in efficient convolutional neural networks. In: CVPR 2019, Computer vision and pattern recognition, Vol. 1. 2019, p. 1–10, <https://arXiv:1910.09455v1> [cs.CV].
- [22] Kingma DP, Ba J. Adam: A method for stochastic optimization. In: International conference on learning representations, Vol. 9. 2015, p. 1–15, <https://arXiv:1412.6980v9> [cs.LG].
- [23] Kandpal I, Castelli M. The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset. *ICT Express* 2020;6(4):312–5. <http://dx.doi.org/10.1016/j.icte.2020.04.010>.
- [24] Irshad H, Kouhsari LM, Waltz G, Bucur O, Nowak JA, Dong F, Knoblauch NW, Beck AH. Crowdsourcing image annotation for nucleus detection and segmentation in computational pathology: evaluating experts, automated methods, and the crowd. In: Pacific symposium on biocomputing (PSB). 2015, p. 294–305. <http://dx.doi.org/10.13140/2.1.4067.0721>.
- [25] Kumar N, Verma R, Sharma S, Bhargava S, Vahadane A, Sethi A. A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE Trans Med Imaging* 2017;36(7):1550–60.

Amit Kumar Chanchal is currently pursuing the Ph.D. degree with the National Institute of Technology Karnataka, Surathkal, India.

Aman Kumar is currently pursuing the B.Tech. degree with the National Institute of Technology Karnataka, Surathkal, India.

Shyam Lal received the Ph.D. degree in image processing from the Birla Institute of Technology, Mesra, Ranchi, Jharkhand, India, in 2013. He is currently working as an Assistant Professor with the Department of Electronics & Communication Engineering, National Institute of Technology Karnataka, Surathkal, Mangalore, India. His research interests include Medical Image Processing and Remote Sensing.

Jyoti R. Kini received the M.D. degree in pathology from the Department of Pathology, Kasturba Medical College at Mangalore, Manipal Academy of Higher Education, Manipal, Karnataka, India, where she is currently an Associate Professor. Her research interests include histopathology, cytopathology, and hematology.