Deep structured residual encoder-decoder network with a novel loss function for nuclei segmentation of kidney and breast histopathology images

Amit Kumar Chanchal¹ · Shyam Lal¹ · Jyoti Kini²

Received: 31 October 2020 / Revised: 10 November 2021 / Accepted: 23 December 2021 /

Published online: 2 February 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

To improve the process of diagnosis and treatment of cancer disease, automatic segmentation of haematoxylin and eosin (H & E) stained cell nuclei from histopathology images is the first step in digital pathology. The proposed deep structured residual encoder-decoder network (DSREDN) focuses on two aspects: first, it effectively utilized residual connections throughout the network and provides a wide and deep encoder-decoder path, which results to capture relevant context and more localized features. Second, vanished boundary of detected nuclei is addressed by proposing an efficient loss function that better train our proposed model and reduces the false prediction which is undesirable especially in healthcare applications. The proposed architecture experimented on three different publicly available H&E stained histopathological datasets namely: (I) Kidney (RCC) (II) Triple Negative Breast Cancer (TNBC) (III) MoNuSeg-2018. We have considered F1-score, Aggregated Jaccard Index (AJI), the total number of parameters, and FLOPs (Floating point operations), which are mostly preferred performance measure metrics for comparison of nuclei segmentation. The evaluated score of nuclei segmentation indicated that the proposed architecture achieved a considerable margin over five state-of-the-art deep learning models on three different histopathology datasets. Visual segmentation results show that the proposed DSREDN model accurately segment the nuclear regions than those of the state-of-the-art methods.

✉ Shyam Lal
shyam.mtec@gmail.com

✉ Jyoti Kini
kinijyoti@gmail.com

Amit Kumar Chanchal
amit.chanchal01@gmail.com

¹ Department of Electronics and Communication Engineering, National Institute of Technology Karnataka, Surathkal, Mangaluru-575025, Karnataka, India

² Department of Pathology, Kasturba Medical College Mangalore, Manipal Academy of Higher Education, Manipal, India

Keywords Kidney cancer diagnosis and prognosis · Nuclei segmentation · Residual learning · Histopathology images

1 Introduction

Recent research trends shows that the deep learning framework performed very well for segmentation, detection, and other computer vision tasks. In the last decade, with the advancement of new types of computation systems, proper strategies to handle overfitting problems to train very deep networks, and lots of changes that are suitable for deep learning networks. Segmentation of haematoxylin and eosin (H & E) from stained histopathology image is the primary prerequisite in artificial pathology. The histopathology slides preparations are discussed by Slaoui M et al. in [27], by the following steps: (I) Tissue collection (II) Fixation (III) Embedding (IV) Sectioning (V) De-paraffining (VI) Staining (VII) Digitizing the slide by whole slide imaging (WSI). There are several tissue collection methods which are fine needle aspiration, biopsy needle, excisional biopsy, etc. A larger biopsy has more information than a small needle biopsy because it preserves the large cellular context of histopathology slides. Fixation of tissue is needed for chemical and physical stabilization. Embedding is required to give a particular shape to the tissue such that it can be easily cut by the machines. Sectioning is required to get all three-dimensional tissue information in the form of many thin slides two-dimensional information. Removing paraffin from the sectioned tissue is important, without de-paraffining the tissue may look a little bit blurry in some of the portions. Staining of the tissue slides is required because it is not visible or kind of transparent under bright field microscopy. The most widely used stains for histopathology images are hematoxylin and eosin. Segmentation tasks can be categorized into traditional or handcrafted feature extraction techniques and CNN-based deep learning approaches. Traditional segmentation methods are mostly based on similarity-based approach, discontinuity-based approach, watershed techniques, active contour methods and their variants, superpixel, and clustering-based methods, etc. The similarity-based approach discussed by Gonzalez R C et al. in [8], is based on local thresholding, global thresholding, adaptive thresholding, Otsu's thresholding, region growing, region splitting, and merging, where these methods try to group and segment similar pixels. For image histogram having flat valleys, the similarity-based approach does not work well and the wrong selection of threshold value may result in over-segmentation and under-segmentation in this case. The discontinuity-based approach tries to segment those pixels which are isolated in some manner like point, line, edges, and it is a mask processing-based approach. This method requires different operators at different stages. Cousty J et al. proposed watershed segmentation method in [4], based on the split, merge, and marker controlled watershed. Detected boundaries in the watershed method depend on cell complexity. Song T et al. proposed active contour segmentation in [28], where they consider intensity information and local edge information for the detection of object boundaries. Superpixel segmentation method used by Albayrak A et al. in [1], is based on the cluster of connected pixels having identical features. It considers the color and coordinate information of neighbor pixels. This technique provides better regional information but is not very effective in the case of cell segmentation. Clustering based segmentation proposed by Win K Y et al. in [37], performs grouping based on their similarity. In recent research work, most of the authors reported that, the segmentation technique based on a deep convolutional neural network performs far better than

the conventional segmentation approach. A concise review of CNN based approach is presented in Section 2. Deep learning segmentation methods also suffer from many challenges. If we categorize these challenges, it will come under the following aspects.

1. Due to large variations of tissue appearance and a varied spectrum of class and sub-class of tissues, it is difficult to recognize.
2. Segmentation of complex boundaries, overlapped boundaries, and vanishing boundaries is not an easy task.
3. Preparation of ground truth in the case of supervised learning is also a big challenge. Supervision of experienced pathologists is necessary since prediction accuracy totally depends on annotated ground truth.

In the case of complex histopathology images, conventional methods suffer either over-segmentation or under-segmentation. The proposed approach focuses to separate the overlapped and vanished nuclear regions from histopathology images. To address the challenges in the segmentation of nuclei from histopathology images our contributions in this paper as follows.

1. To strengthen the multi-level intermediate features, our proposed DSREDN model effectively utilized the strength of residual learning.
2. Through empirical evidence and careful experimentation and analysis, we proposed a novel loss function. Visual results and performance matrices indicate that our loss function better trains the model and accurately segment the nuclear regions compared to the state-of-the-art methods.

The manuscript is organised as follows: A brief introduction about segmentation and challenges associated with it is described in Section 1. A concise review of CNN based approach is presented in Section 2. Sections 3 and 4 cover detailed analysis of the proposed method and their implementation with benchmark datasets. Experimental results and discussions are presented in Section 5 and finally, we concluded our work in Section 6.

2 Related work

Most of the CNN architecture for the cell segmentation task consists of an encoder-decoder path for feature extraction. Many of the recent research utilizes lots of potential opportunities like improving training strategies, handling overfitting problems, better optimization methods, and many other strategies to obtain better prediction accuracy. However, many authors reported their result which is very efficient but an accurate and efficient segmentation algorithm is still open-ended research due to complexity in histopathology images. One of the significant contributions by Ronneberger et al. in [26], called UNet, provides a very good direction and a dramatic breakthrough in the field of biomedical image segmentation. UNet is a symmetric encoder-decoder convolutional network and has a large number of feature channels that allow to extract feature to the higher layer in a deep network. Repeated application of (3 x 3) convolution kernel followed by ReLU activation, (2 x 2) maxpooling and (2 x 2) up-sampling with stride size of 2 and (1 x 1) convolution followed by sigmoid activation at final layer, total of 23 layers in the network. In [36], Veit A et al. realized through their experiment that if a network has a collection of paths then a shorter path is enough during training or a very deep path is not required during training. These multiple paths do not strongly depend on each other and their smooth co-relation

with multiple valid paths increases the performance of the network. In [22], Milletari F et al. proposed an encoder-decoder convolutional network for three-dimensional data by utilizing dice loss as a loss function. Their empirical evaluation achieves better performance on the strong imbalance dataset. In [24], Nogues I et al. proposed architecture for detection of lymph nodes by two fully nested supervised convolutional networks and a structured conditional random field optimization strategy. Degradation of information in deeper network addressed by Kaiming He et al. in [9], by introducing deep residual network which is easier to train and optimize. The residual connection is realized by skipping one or more layers to restore the flow of information in a deep network. For segmentation and detection of histological objects, Chen H et al. in [5], introduced a contour-aware model that extracts multi-level information under auxiliary supervision. In [10], Huang G et al. proposed a convolutional network, which strengthens the overall flow of the input feature map by feeding preceding layer input as well as original input both. Their experiment also indicates that due to integration of identity mapping, model learns more compact features and reduces the vanishing gradient problem. In the case of an imbalanced dataset, predictions are biased towards high precision and low recall which is not tolerable especially in the medical field. This problem is addressed by Salehi S M et al. in [29], which trained the deep network, even with a highly imbalanced dataset, and handled effectively where false negative prediction is much dangerous than false positive. The behavior of loss function such as weighted cross-entropy and dice loss with different learning rates examined by Sudre C H et al. in [30], on medical images and house dataset. Their experiment found that as the level of imbalance increases overlap measure-based loss function is more effective. A very efficient in terms of memory and time for semantic segmentation of road and indoor scene, an encoder-decoder architecture called SegNet by Badrinarayanan V et al. in [3]. SegNet generates a sparse feature decoder that upsamples with the transferred pool and its lower resolution input from its encoder. To accurately segment near boundary regions Zhou S et al. in [38], used a residual network with a dilated convolution block. They utilize many hierarchical blocks in parallel to retrieve meaningful semantic information. To handle class imbalance problems or reducing false-negative predictions in healthcare, Hashemi S R in [11] proposes a 3D-dense CNN with Tversky index-based asymmetric similarity loss that trains the network with the lowest surface distance. Complex boundary-related segmentation problem addressed by Naylor P et al. in [25], by formulating a loss function based on intra-nuclear distance. Their encoder-decoder model outperform on FCN, FCN+PP, Mask R-CNN, U-Net, U-Net+PP experimented with TNBC and MoNuSeg datasets. Meaningful extensions in standard encoder-decoder by incorporating an additional module called attention gate by Schlemper J et al. in [31], and attention as well as residual mechanism by Lal S et al. in [20], where network is trained in such a way that it suppress irrelevant features while highlight the meaningful feature. For road scene segmentation Malekijoo A et al. in [23], utilized the autoencoder-based model where convolution, deconvolution, and pyramid pooling were applied for reinforcing the local feature. For the segmentation of microscopic, MR, and CT images an encoder-decoder architecture by Zhou S et al. in [39], linked meaningful connections to precisely locate the complex boundaries. For the segmentation of nuclei in pathology images, Lal S et al. model [21], consists of adaptive color deconvolution, multi-scale thresholding followed by morphological operations, and other post-processing steps. For the segmentation of medical images, a novel loss function by Karimi D et al. in [16], estimated Hausdorff distance using the morphological operation method, distance transformation method, and by circularly convoluted kernels of different radius. Utilizing methods of reduction of the Hausdorff distance, they train CNN for various microscopy images and

compare their results with a commonly used loss function. Hanif M S et al. in [12], proposed a competitive residual network by stacking multiple residual units called wide network. Their study concluded that the performance of such a wide network is better than the deep and thin network. Chanchal A K et al. and Aatresh A A et al. in [2, 6], used separable convolution pyramid pooling and dimension-wise pyramid pooling for nuclei segmentation tasks. A summary of state-of-the-art DL techniques useful for biomedical image segmentation is presented in Table 1.

3 Proposed architecture

For the segmentation of microscopy images, an encoder-decoder architecture is best suited due to the fact that if an encoder has regular convolution layers, max-pooling layers, then it captures the context in the image very effectively. Decoder path presents the output by gradually applying up-sampling, collecting relevant features from the encoder, and enable precise localization. For each of the filters in the encoder side of the DSREDN network shown in Fig. 1, accepts input of flexible size. we have applied regular (3 x 3) 2D standard convolution, batch normalization, and max-pooling. To avoid the saturation problems and loss of information while going deeper into the network, we restored the lower layered information by creating an additional path parallel to the main path of the network. These two paths are not strongly correlated to each other and it avoids vanishing gradient problems. For each of the filter sizes, the entire encoder side of the DSREDN network consists of three convolution layers in parallel with a single convoluted path that focused to flow the more contextual feature in the network. Since the effectiveness of the decoder path to generate the final output depends on the collection of contextual features from the encoder side, we have a slightly different path on the decoder side, for the optimal processing of the collected feature. By this procedure, our DSREDN network becomes wide and deep instead of thin and deep. DSREDN network trained with RGB images of size (512 x 512 x 3). Five stages of encoder path having five different filter sizes and corresponding decoder path consist of (a) 2D convolution of kernel size (3 x 3) with ReLU activation (b) A high-resolution layer (c) (2 x 2) max-pooling layer in encoder path to reduce the spatial size of the image and corresponding (2 x 2) up-sampling layer in decoder side to collect contextual feature from encoder side by concatenation operation (d) At the final stage a (1 x 1) convolution is used to map the size (512 x 512 x 16) to (512 x 512 x 1) with sigmoid activation.

3.1 Standard convolution layer

Convolution layers have a set of kernels to extract the features and weights of those kernels are automatically learnable. We have given an RGB image of size (512 x 512 x 3). For the input vector $x(j, k)$ and filter size $h(j, k)$ the normal 2D convolution results $y(j, k)$ and is expressed in (1).

$$y(j, k) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x(m, n)h(j - m, k - n) \quad (1)$$

We have an image of (M x M x C), if we apply N kernel of (K x K x C) on it with a step size of S and padding P, the size of resulting image as in (2).

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

Table 1 Summary of state-of-the-art DL techniques used for segmentation of medical images

Author	Application	Dataset	Modalities	Method	Toolbox	Activation Function	Loss Function	Optimizer	Performance Criteria
Ronneberger et al. [26]	Cell segmentation	ISBI- 2012	Light micro-scopic images	2D-CNN, Repeated application of Max-pooling and Up-convolution	Caffe	ReLU and Sigmoid	Cross-entropy	SGD	IOU=92% for PhC-U373 and 77% for DIC-HeLa
Milletari et al. [22]	F Clinical	PROMISE 2012	MRI	Volumetric convolution, Feature channels doubles at each stage	Caffe	PRelu and Softmax	Dice loss	Standard back-propagation	Avg. Dice=86.9%
Nogues et al. [24]	Lymph node cluster segmentation	MICCAI 2015	CT	Holistically-nested neural network, Structured optimization	Caffe	ReLU and Sigmoid	Cross-entropy and Auxiliary loss	SGD	Mean Dice coefficient=82.1% $\pm$ 9.6%
He et al. [9]	Image recognition	ImageNet, CIFAR-10	1000 classes and 10 classes (Colored images)	Deep residual learning, Identity mapping by shortcuts	Caffe	Softmax	Cross-entropy	SGD	Error= 3.57%, 6.43%
Chen et al. [5]	Gland segmentation	Warwick-QU	Histopathology	2D-CNN, Auxiliary supervision, Transfer learning	Caffe	Softmax	Formulated loss on per-pixel classification	Standard back-propagation	F_1 Score=91% and 77%
Huang et al. [10]	Classification of natural images	CIFAR-10, 100, SVHN, and ImageNet	Colored images	Feature-maps of all preceding layers are used as inputs	Keras, Tensorflow	ReLU, Softmax	Categorical crossentropy	SGD	Error=5.19%, 19.64%, 1.59%
Salehi et al. [29]	Training with unbalanced data	MICCAI 2008	MRI	3D fully convolutional network based on the UNet architecture	Nvidia Geforce GTX1080 GPU	ReLU, Softmax	Tversky	Adam	DSC=56.42%, F_2 Score=57.32%
Sudre et al. [30]	Training with unbalanced data	BRATS	MRI	2D and 3D deep learning framework	Tensorflow	ReLU and Sigmoid	Dice	SGD	Dice loss was found to be more robust than the other loss functions

Table 1 (continued)

Author	Application	Dataset	Modalities	Method	Toolbox	Activation Function	Loss Function	Optimizer	Performance Criteria
Badrinarayanan et al. [3]	Indoor and road segmentation	CamVid, SUN RGB-D	RGB	Fully convolutional encoder decoder	Caffe	ReLU and Softmax	Cross-entropy	SGD	Mean IoU=60.10, 31.84
Hashemi et al. [11]	Medical image segmentation	MSSEG-2016, ISBI	MRI	3D fully convolutional, densely connected architecture	Nvidia GeForce GTX1080 GPU	ReLU and Softmax	Asymmetric similarity	Adam	Average Dice=69.9%, 65.74%
Naylor et al. [25]	Segmentation of Nuclei	TNBC, TCGA	Histopathology	Fully convolutional network	Python 3, tensorflow	ReLU and Softmax	Regression loss	Adam	AJL=55.98%, F_1 =78.63%
Schlemper et al. [31]	Medical image analysis	TCIA Pancreas, Multi-class abdominal	CT	Focuses on target structures by employing attention gate	PyTorch	ReLU and Sigmoid	Dice	SGD	Dice score (TCIA)=82% (Multi-class abdominal) =84%
Lal et al. [20]	Liver cancer analysis	KMC liver	Histopathology	Employed residual block, attention mechanism, and joint loss function	TensorFlow 2.0, Keras	ReLU and Sigmoid	Formulated joint loss (Dice and Jaccard)	Adam	F_1 =83.59%, and JI=72.06%
Malekijoo et al. [23]	Indoor and road segmentation	CamVid	RGB	Convolution deconvolution, Pyramid pooling	Python3, Tensorflow 1.4	ReLU and Softmax	Cross entropy	Adam	Mean IoU =48.90%, Accuracy=88.49%
Zhou et al. [39]	Low contrast medical image segmentation	Pelvic CT, Brain tumor, Nuclei segmentation	CT, MR and, Microscopic	Multiscale dense connections, high resolution pathways	Caffe	ReLU and Sigmoid	Difficulty guided cross-entropy	Adam	Dice ratio=95%, and F_1 score= 91.1±10.2

Table 1 (continued)

Author	Application	Dataset	Modalities	Method	Toolbox	Activation Function	Loss Function	Optimizer	Performance Criteria
Lal et al. [21]	Nuclei segmentation	Gold-standard, KMC	Histopathology	Adaptive colour deconvolution, Multi-level thresholding	Python-3	–	–	–	Pr, Re, F_1 = 94.6%, 89.6%, 91.2%, Pr, Re, F_1 = 78.7%, 76.6%, 77.2%
Karimi et al. [16]	Medical image segmentation	TRUS, PROMISE12, 3D CT-Liver, and 3D CT-Pancreas	Ultrasound, MR, and CT	Distance transform, Morphological erosion, Spherical convolution kernels	Python 3.6, TensorFlow 1.2	Softmax	Hausdorff distance	Adam	DSC=95%, 87%, 94%, 78%
Kumar et al. [17]	Clinical	Multiple organs	Histopathology	A CNN that produces a ternary map	PyTorch, NVIDIA Tesla	ReLU and Softmax	Cross entropy	Standard back-propagation	AJL=50.83%, F_1 Score=82.67%
Chanchal et al. [6]	Clinical	TNBC, Kidney, MoNuSeg	Histopathology	Separable convolution pyramid pooling in encoder-decoder network	Keras, TensorFlow, Python-3	ReLU and Sigmoid	BCE	Adam	AJL=70%, 86%, 67%, F_1 Score=82%, 92%, 80%
Aatresh et al. [2]	Clinical	Kidney, TNBC	Histopathology	Attention based encoder-decoder, ASPP, Dimension-wise convolutions	PyTorch	ReLU and Sigmoid	Global	Adam	AJL=87%, 70%, F_1 Score=93%, 82%
Chanchal et al. [7]	Clinical	TNBC, Kidney, MoNuSeg	Histopathology	High-resolution encoder-decoder path, An ASPP at bottleneck	Keras, TensorFlow, Python-3	ReLU and Sigmoid	BCE	Adam	AJL=73%, 94%, 72%, F_1 Score=84%, 97%, 83%

Fig. 1 Proposed deep structured residual encoder-decoder network

For single-step size and no padding the resulting feature size as in (3).

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

3.2 High resolution layer

Proposed DSREDN architecture has a high-resolution encoder and more effective decoder shown in Fig. 2. The mathematical expression of the flow of features is shown below and it can be seen that these two paths are not strongly dependent on each other and their smooth co-relation with multiple valid paths increases the performance of the network and will reduce the vanishing gradient problem. This wide network strengthens the overall extracted feature map by feeding preceding layer input as well as the original input. Our experiment also indicates that due to the integration of this wide network our model got the additional discriminative capability and is able to retrieve more compact features compared to the other existing deep model.

Aggregation of features for high-resolution encoder path is as follows:

$$\begin{aligned} X_3 &= F_3(X_2) + X_1 \\ &= F_3(F_2(X_1)) + F_1(X_0) \\ &= F_3(F_2(F_1(X_0))) + F_1(X_0) \end{aligned}$$

Aggregation of features for high resolution decoder path can be expressed as:

$$\begin{aligned} X_3 &= F_3(X_2) + X_{11} \\ &= F_3(F_2(X_1) + X_{11}) + X_{11} \\ &= F_3(F_2(F_1(X_0)) + F_{11}(X_0)) + F_{11}(X_0) \end{aligned}$$

Fig. 2 High resolution encoder path (Left) and decoder path (Right)

3.3 Activation function

If the dataset is linearly separable, then the linear activation function does well, but in the real world, the dataset is rarely linearly separable. The output of the neural network is based on the linear operation of variables and a linear function does not allow the model to learn complex relations. Nonlinear activation function can process almost any nonlinear relation and provide very good prediction results. Nonlinearity helps the model to adapt or generalized with a variety of data. Rectified Linear Unit (ReLU) is the most popular activation function in deep learning models. We have used ReLU activation function in all of the hidden layers of the DSREDN model due to computational simplicity, representational sparsity, and its linear behavior. ReLU activation function and its derivative are expressed mathematically by (4), and (5). At the final stage a (1 x 1) convolution is used to map the size (512 x 512 x 16) to (512 x 512 x 1) with sigmoid activation.

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

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

3.4 Pooling layer

To make the architecture scale-invariant, rotation invariant, and location invariant pooling is used. pooling operation on (4 x 4) with a step size of S= 2 and kernel K=(2 x 2) is shown in Fig. 3.

3.5 Batch normalization

A deep network has large number of layers and there are lots of activity happening between these layers. It has been observed that if input at each of the pre-activation is with a single distribution, ideally it should be Gaussian distributed or mean-centered then it is easier to train a very deep network. A constant and small change at an earlier stage of input leads to a significant effect on the latter layer and in that case due to internal co-variate shift shown in Fig. 4, it is difficult to train a very deep network. To avoid such problems we have applied batch-normalization used by Ioffe S et al. in [13], so that all input coming from the

Fig. 3 Max Pooling, step size S=2 kernel K=2

Fig. 4 Internal co-variate shift of batches

previous layer guaranteed to be the same distribution and convergence become faster. Batch normalization is an additional layer and works as a regularizer.

4 Training and implementation

We have used Google Colab notebook (with 12 GB Google GPU), Tensorflow 2.0, and Keras API for simulation of all models. The most important part of any optimization method in deep learning is about gradient which is used to update the weights. Equation (6) is a general weight update equation.

$$w_t = w_{t-1} - \eta \left[\frac{\partial L}{\partial w} \right] \quad (6)$$

We have used Adam (Adaptive Moment Estimation) optimizer. Adam [19] has integrated with nice features of the RMSProp and Stochastic gradient descent (SGD) with Momentum (SGD +Momentum) algorithms. The final update equation is a combination of RMSProp and SGD with Momentum. Adam is one of the fastest optimization approaches for recent research trends. Followings are the significant controlling parameters that we fine-tuned: (a) Learning rate= 0.0001 for training of all the model with three datasets (b) Weight decay constant, we have selected learning rate and weight decay constant based on trial and error (c) Batch size=4, Larger batch size allow us to parallelize computations to a greater degree but lead to poor generalization (D) The size of the convolution filter used as per available resources. We have considered F1-score used by Lal S et al. and Aatresh A A et al. in [2, 21], and Aggregated Jaccard Index (AJI) used by Naylor P et al. in [25], the total number of parameters, and FLOPs (Floating point operations), which are mostly preferred performance measure metrics for comparison of nuclei segmentation.

4.1 Datasets

- (I) Kidney dataset: Kidney dataset was prepared by Irshad H et al. [14]. it consists of 730 H&E images (400 Pixels x 400 Pixels) and their corresponding ground truth.
- (II) Triple negative breast cancer (TNBC) dataset: TNBC dataset was prepared by Naylor P et al. [25]. This dataset consists of 50 H&E images (512 Pixels x 512 Pixels) of

breast tissue. We performed data augmentation like horizontal flip, vertical flip and rotation such that we have sufficient number of training images.

- (III) MoNuSeg-2018 dataset: This dataset was prepared by Kumar N et al. [17] and included seven different organs of colon, breast, kidney, liver, stomach, prostate, and bladder. It consists of 44 images (1000 Pixels x 1000 Pixels). We performed patch processing of obtained images to make the input image of dimension (512 Pixels x 512 Pixels) and some data augmentation techniques to make the training set larger.

4.2 Loss function

The deep learning model learns by means of loss function which is driven by a calculation of error. This error is the difference of actual value and predicted value as in the regression problem and the difference of actual distribution and predicted distribution in the case of a classification problem. With the help of suitable optimization algorithm, the loss function learns to reduce the error. Following are the most preferred loss function in the case of binary segmentation.

4.2.1 Binary cross-entropy

Binary Cross-Entropy is a combination of sigmoid activation and Cross-Entropy, which is discussed by Chanchal et al. in [6]. we have a countable set of symbols $X = \{x_1, x_2, \dots, x_i, x_n\}$. Y is the discrete probability distribution which represents the probability of occurrence of those symbols. $Y = \{y_1, y_2, \dots, y_i, y_n\}$. Let $y_i = p(x_i)$ where y_i is the probability of occurrence of symbol x_i . According to the concept of entropy the minimum number of bits required to represent the i^{th} symbol is $x_i = \log(1/y_i)$. If we consider entire distribution Y to achieve the optimal number of bits per transmission through some channel, then the optimal number of bits is known as entropy. Mathematically it is just the expected number of bits per encoding and can be shown in (7).

$$H(Y) = \sum_{i=1}^n y_i \log \frac{1}{y_i} = - \sum_{i=1}^n y_i \log y_i \quad (7)$$

now let $X' = \{x'_1, x'_2, \dots, x'_i, x'_n\}$ with probability of occurrence $Y' = \{y'_1, y'_2, \dots, y'_i, y'_n\}$ if we encode symbol X using different symbol X' then encoding will require $x'_i = \log \frac{1}{y'_i}$ instead of $\log \frac{1}{y_i}$ then we define Cross-Entropy $H(Y, Y') = \sum_{i=1}^n y_i \log \left(\frac{1}{y'_i} \right)$. If distributions of Y and Y' are equal then entropy and cross-entropy are equal. For binary cross-entropy we consider two classes, class C_1 and class C_2 . (a) $Label_1 \{0, 1\}$ represents the ground truth label for class C_1 , S_1 represents the sigmoid score for class C_1 (b) $Label_2 = \{1 - Label_1\}$ represents the ground truth label for class C_2 , $S_2 = (1 - S_1)$ represents the sigmoid score for class C_2 . Binary cross-entropy (BCE) used by Ronneberger et al. in [26] and is defined by (8), and (9).

$$S_i = \frac{1}{1 + e^{-(Prediction)_i}} \quad (8)$$

$$BCE = -t_1 \log(f(s_1)) + (1 - t_1) \log(1 - f(s_1)) \quad (9)$$

4.2.2 Weighted binary cross-entropy

For skewed data sometimes weighted binary cross entropy (WBCE) loss function performs better. Normally for minority class we assign higher weights. The purpose of using class weights is to change the loss function so that the training loss cannot be minimized within a limit. It is a way of passing weights to the binary cross entropy loss function used by Jadon S et al. and Sugino T et al. in [15, 35]. Equation (10) describes the weighted binary cross entropy loss function. Here β is used to assign weights to the more relevant objects. y , and $\hat{y}$ represents ground truth and predicted results respectively.

$$L_{\text{Weighted BCE}}(y, \hat{y}) = -\beta * y \log(\hat{y}) - (1 - y) \log(1 - \hat{y}) \quad (10)$$

4.2.3 Dice loss

In case of binary segmentation and having class imbalance dice loss is a most preferred overlap measure used by Milletari F et al. and Sudre C H et al. in [22, 30]. There is little difference in the calculation of loss using intersection over union (IOU) and dice. Equations (11), (12), and (13) describes the dice loss function.

$$IOU = \frac{\text{Intersection}}{\text{Union}} = \frac{TP}{FP + TP + FN} = \frac{\sum \hat{y} \cdot y}{\sum \hat{y} + y - \sum \hat{y} \cdot y} \quad (11)$$

While the calculation of dice is by calculating the harmonic mean of precision and recall.

$$DICE = \frac{2 * \text{Intersection}}{\text{Union} + \text{Intersection}} = \frac{2 * TP}{FP + 2 * TP + FN} = \frac{2 * \sum \hat{y} \cdot y}{\sum \hat{y} + y} \quad (12)$$

$$Loss = 1 - \frac{2 * \sum \hat{y} \cdot y}{\sum \hat{y} + y} \quad (13)$$

4.2.4 Proposed loss function

The proposed loss function has conceived the idea of dynamically scaled cross-entropy loss. Dynamically scaled cross entropy loss which is a pixel based loss that automatically assigns more weight to the portion that of our interest or object portion and down-weight to those which are of less interest with the help of hyper parameter α_t and γ coefficient.

Dynamically scaled cross-entropy loss For highly imbalanced dataset dynamically scaled cross-entropy loss predict with better accuracy. It is a dynamically scaled cross entropy loss that automatically assigns more weight to the portion that of our interest or object portion and down-weight to those which are of less interest. From BCE loss we can derived dynamically scaled cross-entropy loss Jadon S et al. and Sugino T et al. in [15, 35], and can be expressed in (14), (15), (16), and (17).

$$BCE \text{ Loss} = \begin{cases} -\log(\hat{y}) & \text{for } y = 1 \\ -\log(1 - \hat{y}) & \text{otherwise} \end{cases} \quad (14)$$

To make the notation more convenient we can write:

$$p_t = \begin{cases} \hat{y} & \text{for } y = 1 \\ 1 - \hat{y} & \text{otherwise} \end{cases} \quad (15)$$

Table 2 Superiority of proposed loss function **F1/A_{JI}** with benchmark model

Dataset Model	(Kidney) UNet	(TNBC) UNet	(MoNuSeg) UNet	(Kidney) Proposed	(TNBC) Proposed	(MoNuSeg) Proposed
BCE	0.8537/0.7489	0.7324/0.6559	0.7278/0.6029	0.9422/0.8994	0.8008/0.6698	0.7854/0.6675
Dice	0.8699/0.7631	0.7470/0.6690	0.7186/0.5991	0.9479/0.9044	0.8100/0.6828	0.7927/0.6607
WBCE	0.8408/0.7376	0.7360/0.6585	0.7168/0.5938	0.9460/0.9115	0.7930/0.6588	0.7919/0.6597
Focal	0.8622/0.7563	0.7228/0.6467	0.7399/0.6088	0.9346/0.8990	0.8175/0.6928	0.7940/0.6619
Tvesky	0.8793/0.7713	0.7573/0.6782	0.7423/0.6149	0.9286/0.9040	0.8346/0.7182	0.7810/0.6513
BCE+Dice	0.8479/0.7329	0.7250/0.6493	0.7321/0.6065	0.9472/0.9010	0.8087/0.6796	0.7950/0.6637
Proposed Loss	0.8878/0.7732	0.7653/0.6854	0.7460/0.6179	0.9579/0.9261	0.8517/0.7415	0.8065/0.6795

Values in bold indicate the highest performance score of the proposed loss function among other comparable loss functions

Now BCE can be written in a very precise manner as follows:

$$BCE(y, \hat{y}) = BCE(p_t) = -\log(p_t) \quad (16)$$

A modulation factor $(1 - p_t)^\gamma$, $\gamma > 0$ to down-weight the background and assign more weight to object region and a hyper-parameter α_t , $0 < \alpha < 1$ make the above loss function dynamically scaled cross-entropy loss.

$$\text{Dynamically Scaled Cross - Entropy Loss } (p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (17)$$

Fig. 5 Accuracy and loss plots (Kidney Dataset) (a) Accuracy plot with proposed loss function (b) Accuracy plot with BCE loss function (c) Loss plot with proposed loss function (d) Loss plot with BCE loss function

Approximation of dice loss This loss derived from dice coefficient by setting different weight to false positive and false negative with the help of β coefficient. Mathematically This loss can be expressed in (18).

$$\text{Approximated Dice Loss} = (1 - DI) \text{ where } DI = \frac{\sum \hat{y} \cdot y}{\sum \hat{y} \cdot y + \beta(1 - y)\hat{y} + (1 - \beta)y(1 - \hat{y})} \quad (18)$$

Proposed loss function calculated based on the concept of dynamically scaled cross entropy loss in (17). Segmentation results indicated that the dynamically scaled Dice loss shown in (19) better train any encoder-decoder model with histopathological data. Superiority of proposed loss function with F1-score/IOU score of two benchmark model is shown in Table 2.

$$\text{Dynamically Scaled Dice Loss} = \alpha_t(1 - DI)^\gamma \quad (19)$$

DI indicates Dice index and the value of α_t and γ ranges $0 < \alpha_t < 1$, $1 < \gamma < 3$, β value can be used to tune false negative and false positive.

The training and validation curve showed in Figs. 5, 6, and 7 indicates the proposed loss function better represents the validation data. For the best representation of the model,

Fig. 6 Accuracy and loss plot (TNBC Dataset) (a) Accuracy plot with proposed loss function (b) Accuracy plot with BCE loss function (c) loss plot with proposed loss function (d) Loss plot with BCE loss function

Fig. 7 Accuracy and loss plot (MoNuSeg Dataset) (a) accuracy plot with proposed loss function (b) accuracy plot with BCE loss function (c) loss plot with proposed loss function (d) loss plot with BCE loss function

the training and validation curve should be closer to each other and for optimal bias and variance, they should be collapsed to each other. The model has both curves closer to each other, that model is robust during testing. The large gap between the training and validation curve indicates the model is working very well in training data but not as much as for validation data. Using proposed loss our model is much generalized to work on different types of data.

Table 3 Performance comparison of different models with three datasets

Model	F1 AJI		F1 AJI		F1 AJI	
	(Kidney Dataset)		(TNBC Dataset)		(MoNuSeg Dataset)	
Unet [26] (2015)	0.8537	0.7489	0.7324	0.6559	0.7278	0.6029
SegNet [3] (2017)	0.8972	0.8304	0.7685	0.6434	0.7879	0.6514
Dist [25] (2019)	0.8992	0.8272	0.7516	0.6727	0.7795	0.6467
Att. UNet [31] (2019)	0.9135	0.8590	0.7216	0.6194	0.7735	0.6315
HMEDN [39] (2020)	0.9262	0.8676	0.8124	0.7017	0.7865	0.6522
Proposed Model	0.9579	0.9261	0.8517	0.7415	0.8065	0.6795

Values in bold indicate the highest performance score of the proposed model among other comparable models

Table 4 Computational complexity comparison of different models

Model	Parameters(millions)	FLOPs(millions)
Unet [26] (2015)	31.3	62.7
SegNet [3] (2017)	18.8	47
Dist [25] (2019)	7.7	15.5
Att. UNet [31] (2019)	31.9	63.7
HMEDN [39] (2020)	0.20	0.45
Proposed Model	4.3	8.17

Values in bold indicate the highest performance score of the proposed model among other comparable models

5 Results and discussion

We compared our proposed model with five other CNN models which is a benchmark in the field of biomedical image segmentation. We expressed our simulation results in terms of F1-Score used by Lal S et al. and Aatresh A A et al. in [2, 21] and Aggregated Jaccard Index (AJI) used by Naylor P et al. in [25]. By calculating the harmonic mean of precision and recall, the F1 score is calculated and is the most preferred method to measure the retrieved information. AJI is a connected component-based method which is the improved version of the pixel-based global Jaccard Index. A higher AJI score indicates a better segmentation model. In Table 3, we compared the DSREDN model with five benchmark models for three histopathological datasets, and Table 4 shows a computational complexity comparison of the proposed DSREDN architecture with other segmentation architecture. Performance is measured in terms of F1 Score, AJI score, and the total number of trainable parameters that describe the training time and complexity. We have considered six sample test images, (two from each of the histopathological datasets) for visualization of results, and presented in Figs. 8, 9, and 10.

5.1 U-Net prediction

U-Net architecture has repeated application of 3x3 unpadded convolution, a total of 23 convolution layers, 2x2 max-pooling for downsampling, 2x2 upsampling, and each convolution layer followed by ReLU activation. Finally, 1x1 convolution is followed by sigmoid activation. U-Net architecture clearly identifies 51 nuclei out of 57 in image 1, 45 nuclei out of 53 nuclei in image 2. Around 10% to 15% nuclei are not clearly identified by this architecture. Some of the nuclei are in clustered form and some are not predicted. Two nuclei are clustered in image-1 and six are in image 2. Four additional ducts and six additional ducts are predicted in image 1 and image-2 respectively which are actually not desirable. This architecture almost follows a similar prediction pattern in the other four images.

5.2 SegNet prediction

A very efficient in terms of memory and time for semantic segmentation of road and indoor scene an encoder-decoder architecture called SegNet. To generate a sparse feature, decoder upsample with the transferred pool and its lower resolution input from its encoder. Its encoder has 13 convolution layers, followed by ReLU activation similar to VGG-16, batch normalized, 2x2 max-pooling, and corresponding 13 layer decoder. SegNet has a slightly

Fig. 8 Comparison of predicted nuclear regions of five state-of-the-art models on kidney images

different decoder that uses the max-pooling indices to upsample the feature and convolution with a trainable filter bank. Visual segmentation by SegNet indicates that a number of clearly detected nuclei is (80–85%), which is less than U-net. Overlapped nuclei are lesser than U-Net and no undesirable things are detected by this model. The number of partially detected nuclei increases compared to U-Net and is three and two in image-1 and image-2 respectively. The number of nuclei not identified is maximum in this case.

5.3 Attention U-Net prediction

Meaningful extensions in standard U-Net by incorporating an additional unit called attention gate that is trained in such a way that it suppresses irrelevant features while highlighting the meaningful feature by strengthening the capability of the decoder. With the use of attention coefficients, this architecture increases the receptive field which is the key for better

Fig. 9 Comparison of predicted nuclear regions of five state-of-the-art models on TNBC images

semantic segmentation. Their attention module can be easily integrated into any other segmentation method. The number of clearly identified nuclei is almost similar to the SegNet model but partially detected nuclei is maximum in this case. Almost 5% of the nuclei are not predicted by this model.

5.4 Dist prediction

Complex boundary-related segmentation problem addressed by formulating a loss function based on intranuclear distance. They compare their model, namely DIST with FCN, FCN+PP, Mask R-CNN, U-Net, U-Net+PP for triple-negative breast cancer dataset and other datasets which are of seven different organs. In this architecture, all nuclei are identified either partially or completely. 82–88% nuclei clearly detected. The problem of segmentation of clustered nuclei, somehow solved, but not completely solved. Partially detected nuclei are less as compared to SegNet, U-Net, and Attention U-Net.

Fig. 10 Comparison of predicted nuclear regions of five state-of-the-art models on MoNuSeg images

5.5 HMEDN prediction

For the segmentation of microscopic, MR, and CT images an encoder-decoder architecture that linked meaningful connections to precisely locate the complex boundaries. A dense connection path with dilated convolution blocks guided by modified binary cross-entropy, accurately detects vanishing boundaries of blurry images. In this architecture, prediction accuracy is slightly improved as compared to the Dist algorithm, but with an improvement in accuracy, false detection cases also increased in HMEDN. Four additional things in image-1 and two undesirable things in image-2 have been found.

5.6 Proposed DSREDN prediction

The best part of our model is that 90% of the nuclei are clearly identified and the rest of the nuclei are either in clustered form or partially detected. Out of 57 and 53 nuclei in image-1

and image-2, 55 and 48 nuclei clearly detected. Less number of additional and undesirable ducts were detected in the proposed model. Partially detected nuclei are very less in number. The predicted image has a morphology closer to the ground truth image. The problem with overlapped nuclei is somehow solved, but not completely solved.

5.7 Clinical significance

Less number of undesirable nuclei indicates the less number of false-positive cases, that are less in number in the proposed model. The number of nuclei that have not been detected is a case of false-negative, which is never desirable especially in health care. The predicted morphology closer to the ground truth image means those images have high clinical and diagnostic values. For clinical purposes, the proposed DSREDN model outperforms the five most recent state-of-the-art models.

5.8 Limitations

Reported results of different datasets show that there is a lack of generalizability in the segmentation of the nuclear region from histopathology images. This is mainly due to the histopathology slide and their corresponding ground truth preparation since the clinical significance of predicted output is highly dependent on the prepared slide and their semantic pixel-wise labeling. Another issue of this study is that segmented boundaries are not fine enough and it is still sub-optimal for clinical use. Issue of partially detected nuclei, overlapped nuclei, and false-positive cases found to be proportional to cell complexity.

6 Conclusion

This paper proposed a CNN-based architecture called a deep structured residual encoder-decoder network (DSREDN), that addressed two major concerns in automatic nuclei segmentation. The first major concern was to identify nuclei from histopathology images having a widely varied spectrum with a large number of artifacts. This problem was addressed by introducing a powerful encoder-decoder having two paths that have more discriminative capability and were able to retrieve relevant and compact textural information. 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. We proposed an efficient loss function through careful experimentation and analysis to segment the nuclei having complex or vanishing boundaries which were the second major issue in the segmentation task. We have used the most preferred performance matrices F1-score and AJI score by performing experiments on the three different publicly available H&E stained histopathological datasets. The obtained quality metrics and predicted nuclear regions of the proposed framework were better in comparison to those of the state-of-the-art models.

Although the proposed model produced excellent results, the feature space may be enriched further by incorporating a high-performance feature extraction module. Also, the proposed method can be generalized to work on more image modalities. This study is a binary segmentation of histopathology images, here we can only segment the nuclear regions. In future, we can grade these nuclear regions into their sub-types. Few innovative applications of different image modalities were reported by Shoeibi A et al. in [32, 33], in which generative adversarial networks (GANs), recurrent neural networks (RNNs),

autoencoders (AEs), convolutional neural networks (CNNs), deep neural networks (DNNs), and other hybrid networks have been developed for automated detection of COVID-19 and multiple sclerosis. In [18, 34], Khodatars M et al. and Sadeghi D et al. illustrated the applicability of deep learning for the diagnosis of autism spectrum disorder and schizophrenia disease detection. These examples highlight how the field of computer-aided diagnosis systems is changing rapidly, and that there may still be numerous applications that have not been focused on yet.

Acknowledgements 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 Interests Authors declare that they have no conflict of interest.

References

1. 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
2. Aatresh AA, Yatgiri RP, Chanchal AK, Kumar A, Ravi A, Das D, Raghavendra BS, Lal S, Kini J (2021) Efficient deep learning architecture with dimension-wise pyramid pooling for nuclei segmentation of histopathology images. *Comput Med Imaging Graph* 93:101975. <https://doi.org/10.1016/j.compmedimag.2021.101975>
3. 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
4. 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
5. 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 [cs.CV]
6. 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. <https://doi.org/10.1016/j.compeleceng.2021.107177>
7. Chanchal AK, Lal S, Kini J (2021) High-resolution deep transferred ASPPU-Net for nuclei segmentation of histopathology images. *Int J Comput Assist Radiol Surg*. <https://doi.org/10.1007/s11548-021-02497-9>
8. Gonzalez RC, Woods RE (2006) Digital image processing, 3rd edn. Prentice Hall, New York, USA. ISBN-013168728X
9. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. 2016 IEEE conference on computer vision and pattern recognition (CVPR), Las Vegas, NV. pp 770–778. <https://doi.org/10.1109/CVPR.2016.90>
10. Huang G, Liu Z, Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. IEEE conference on computer vision and pattern recognition (CVPR), Honolulu. pp 2261–2269. <https://doi.org/10.1109/CVPR.2017.243>
11. Hashemi SR, Salehi SM, Erdogmus D, Prabhu SP, Warfield SK, Gholipour A (2019) Asymmetric loss functions and deep densely-connected networks for highly-imbalanced medical image segmentation: application to multiple sclerosis lesion detection. In: IEEE Access, vol 7, pp 1721–1735. <https://doi.org/10.1109/ACCESS.2018.2886371>
12. Hanif MS, Bilal M (2020) Competitive residual neural network for image classification. *ICT Express* 6(1):28–37. <https://doi.org/10.1016/j.ict.2019.06.001>
13. Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. *Machine Learning*. arXiv:1502.03167
14. 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), pp 294–305. <https://doi.org/10.13140/2.1.4067.0721>

15. Jadon S (2020) A survey of loss functions for semantic segmentation. [Online]. Available: [arXiv:2006.14822](https://arxiv.org/abs/2006.14822)
16. Karimi D, Salcudean SE (2020) Reducing the Hausdorff distance in medical image segmentation with convolutional neural networks. *IEEE Trans Med Imaging* 39(2):499–513
17. 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
18. Khodatars M, Shoeibi A, Sadeghi D, Ghaasem N, Jafari M, Moridian P, Khadem A, Alizadehsani R, Zare A, Kong Y, Khosravi A, Nahavandi S, Hussain S, Acharya UR, Berk M (2021) Deep learning for neuroimaging-based diagnosis and rehabilitation of Autism Spectrum Disorder: A review. *Comput Biol Med* 139:104949. <https://doi.org/10.1016/j.compbimed.2021.104949>
19. Kingma DP, Ba J (2015) Adam: A method for stochastic optimization. In: International conference on learning representations, vol 9. arXiv:1412.6980v9 [cs.LG]
20. Lal S, Das D, Alabhya K, Kanfode 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
21. Lal S, Kanfode A, Alabhya K, Dsouza R, Kumar A, Chanchal AK, Maneesh M, Peryail G, Kini J (2020) A robust method for nuclei segmentation of H&E stained histopathology images. 7th IEEE international conference on signal processing and integrated networks (SPIN2020), Amity University Delhi NCR, Noida, UP
22. Milletari F, Navab N, Ahmadi SA (2016) V-Net: fully convolutional neural networks for volumetric medical image segmentation, fourth international conference on 3D vision (3DV). Stanford, CA, pp 565–571. <https://doi.org/10.1109/3DV.2016.79>
23. Malekijoo A, Fadaeieslam MJ (2019) Convolution-deconvolution architecture with the pyramid pooling module for semantic segmentation. *Multimed Tools Appl* 78:32379–32392. <https://doi.org/10.1007/s11042-019-07990-7>
24. Nogue I et al (2016) Automatic lymph node cluster segmentation using holistically-nested neural networks and structured optimization in CT images. In: Medical image computing and computer-assisted intervention – MICCAI 2016. Lecture notes in computer science, vol 9901. Springer, Cham. https://doi.org/10.1007/978-3-319-46723-8_45
25. 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
26. Ronneberger O, Fischer P, Brox T (2015) U-Net: Convolutional networks for biomedical image segmentation. In: Proc. MICCAI. Springer, Munich, Germany, pp 234–241
27. Slaoui M, Fiette L (2011) Histopathology procedures: from tissue sampling to histopathological evaluation. *Methods Mol Biol (Methods Protoc)* 691:69–82
28. Song T, Sanchez V, ElDaly 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
29. Salehi SM, Erdogmus D, Gholipour A (2017) Tversky loss function for image segmentation using 3D fully convolutional deep networks. In: Proc, Int Workshop Mach Learn Med Imag. Springer, Cham, Switzerland, pp 379–387
30. Sudre CH, Li W, Vercauteren T, Ourselin S, Cardoso MJ (2017) Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer, pp 240–248
31. 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(ISSN 1361-8415):197–207
32. Shoeibi A, Khodatars M, Alizadehsani R, Ghassemi N, Jafari M, Moridian P, Khadem A, Sadeghi D, Hussain S, Zare A, Sani ZA, Bazeli J, Khozeimeh F, Khosravi A, Nahavandi S, Acharya UR, Shi P (2020) Automated detection and forecasting of covid-19 using deep learning techniques: a review. *Machine Learning*. arXiv:2007.10785 [cs.LG]
33. Shoeibi A, Khodatars M, Jafari M, Moridian P, Rezaei M, Alizadehsani R, Khozeimeh F, Gorriz JM, Heras J, Panahiazar M (2021) Applications of deep learning techniques for automated multiple sclerosis detection using magnetic resonance imaging: A review, Image and Video Processing. arXiv:2105.04881
34. Sadeghi D, Shoeibi A, Ghassemi N, Moridian P, Khadem A, Alizadehsani R, Teshnehlav M, Gorriz JM, Nahavandi S (2021) An overview on artificial intelligence techniques for diagnosis of schizophrenia based on magnetic resonance imaging modalities: methods, challenges, and future works. *Machine Learning*. arXiv:2103.03081

35. Sugino T, Kawase T et al (2021) Loss weightings for improving imbalanced brain structure segmentation using fully convolutional networks, healthcare. MDPI 9(8):938
36. Veit A, Wilber M, Belongie S (2016) Residual networks behave like ensembles of relatively shallow networks. Neural Inf Process Syst pp 550–558. arXiv:1605.06431
37. Win KY, Choomchuay S, Hamamoto K (2017) K mean clustering based automated segmentation of overlapping cell nuclei in pleural effusion cytology images. International conference on advanced technologies for communications (ATC). pp 265–269. <https://doi.org/10.1109/ATC.2017.8167630>
38. Zhou S, Nie D, Adeli E, Gao Y, Wang L, Yin J, Shen D (2018) Fine-grained segmentation using hierarchical dilated neural networks. In: Medical image computing and computer assisted intervention, vol 11073. Springer, Cham, pp 488–496
39. Zhou S, Nie D, Adeli E, Yin J, Lian J, Shen D (2020). In: IEEE Transactions on Image Processing, vol 29, pp 461–475. <https://doi.org/10.1109/TIP.2019.2919937>

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