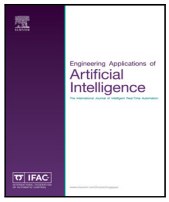
Engineering Applications of Artificial Intelligence 118 (2023) 105704

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai



Saliency and ballness driven deep learning framework for cell segmentation in bright field microscopic images

S.B. Asha [a], G. Gopakumar [a,*], Gorthi R.K. Sai Subrahmanyam [b]

[a] Department of Computer Science and Engineering, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Amritapuri, India
[b] Department of Electrical Engineering, Indian Institute of Technology Tirupati, Tirupati, India

ARTICLE INFO

Keywords:
Deep learning based semantic segmentation
Encoder–decoder model
Microscopic image analysis
Bright-field microscopy

ABSTRACT

Cell segmentation is the most significant task in microscopic image analysis as it facilitates differential cell counting and analysis of sub-cellular structures for diagnosing cytopathological diseases. Bright-field microscopy is considered the gold standard among different types of optical microscopes used for cell analysis due to its simplicity and cost-effectiveness. However, automatic cell segmentation in bright field microscopy is challenging due to imaging artifacts, poor contrast, overlapping cells, and wide variability of cells. Also, the availability of labeled bright-field images is limited, further constraining the research in developing supervised models for automated cell segmentation. In this research, we propose a novel cell segmentation framework termed Saliency and Ballness driven U-shaped Network (SBU-net) to overcome these challenges. The proposed architecture comprises a novel data-driven feature fusion module that enhances the perceivable structure of cells using its saliency and ballness features. This, together with an encoder–decoder model having dilated convolutions and a novel combination loss function, captured the global context of cell structures and produced accurate cell segmentation results. SBU-net is evaluated using two publicly available bright-field datasets of T cells and pancreatic cancer cells. The model is subjected to 5-fold cross-validation and outperformed state-of-the-art models by producing mean Intersection over Union (IoU) scores of 0.804, 0.829, and mean Dice of 0.891, 0.906, respectively. The architecture was also tested on a fluorescent dataset to see how well it could generalize, and it came out with a mean IoU of 0.892 and a mean Dice of 0.948, outperforming other models reported in the literature.

1. Introduction

Cell image analysis is integral in diagnosing, treating, and monitoring patients with various health conditions, and microscopy remains the gold standard for cell analysis. Microscopic images of blood smears, tissue samples, and body fluid samples are analyzed to obtain quantitative information on cell morphology. Cell signature is determined by the shape and size of the cells, morphology of the nucleus, presence of granules, and amount of cytoplasm. These factors may vary depending upon the disease, and microscopic image analysis can be used to assess the variations. Accurate quantification of the cell signature depends on detecting the spatial locations of the cell, and cellular structures in the image (Dimopoulos et al., 2014). It plays a vital role in the detection, treatment, and monitoring of anemias, malaria parasites (Gopakumar et al., 2018), tuberculosis (Simon et al., 2019), eosinophilia, and different types of cancers, including leukemia (Kalmady et al., 2017; Thanmayi et al., 2021).

Traditionally, medical experts manually extract relevant information from microscopy images, which is tedious and time-consuming. Furthermore, the results mainly depend on the technical expertise of the examiner and variability across capturing instruments. Bright-field microscopy is the simplest and cheapest optical microscopy technique used in resource-limited clinics where cost and human expertise are of primary concern (Mualla et al., 2018). Despite its advantages, bright-field microscopy typically exhibits low contrast due to the variation in light absorption by the various biological samples. As a result, analysis of bright-field images remains a challenge. Semi-automated or fully automated systems are required to abstract specific information from the microscopic images in a short amount of time. Although fully automated systems cannot be considered to replace the intelligence of medical professionals, they can facilitate faster and more accurate decisions on case variants (Shaukat et al., 2020). Therefore, one of the focus areas in the analysis of microscopy images is the automated detection and segmentation of cells, and cellular structures (Song et al., 2018). This will help to quantify the severity of many diseases, including malaria and anemia. It also helps in differential cell counting to assess an individual's health.

* Corresponding author.
E-mail address: gopakumarg@am.amrita.edu (G. Gopakumar).

https://doi.org/10.1016/j.engappai.2022.105704
Received 20 February 2022; Received in revised form 8 November 2022; Accepted 30 November 2022
Available online 9 December 2022

Over the past few years, advances in deep learning techniques have outweighed the classical image analysis techniques in the cell and nuclei segmentation (Xing et al., 2018). It is not an easy task due to challenges such as the heterogeneous shape of cells in the image, intracellular variability, and the occurrence of cells as a cluster. The imaging artifacts, overlapping and touching nuclei and cells, and the appearance of cells and nuclei as dense regions are still some of the open problems in automated image segmentation. Besides, the availability of publicly accessible annotated data that can be used to learn the model is insufficient (Dimopoulos et al., 2014). Furthermore, the performance of segmentation algorithms is significantly affected due to the low contrast property of bright-field images.

Several studies based on deep learning techniques (Ali et al., 2021; Moen et al., 2019) have shown promising results for bright-field images, but there is still a wide gap in the performance compared to fluorescence images. Therefore, there is still a great demand for precise, standardized, and robust whole cell segmentation algorithms for measuring the properties and subcellular structures in cell images, especially in bright-field microscopy images (Al-Kofahi et al., 2018). Owing to the simplicity and cost-effectiveness of bright field microscopy, developing an accurate cell segmentation framework will be a beneficial but challenging task. In this research, we propose a neural network architecture that uses a novel loss function to segment out cells accurately, even if there are overlapping cells in the image. Unlike other prevailing segmentation models, the proposed model learns deep features from original images and their corresponding perceptual feature maps. The perceptual features are derived from the data without any manual effort, and a set of specific features, namely, saliency, ballness, and orientation, are selected from those using a voting mechanism. These act as prior information and are used to produce feature-enhanced images for the proposed model. We have shown that the proposed framework outperformed the state-of-the-art models in semantic cell segmentation ability in Section 4(Table 9).

The main contributions of this research are listed below and are discussed in detail in Section 3.

1. A novel feature fusion approach that combines the power of deep learning-based segmentation networks and perceptual features derived from the images is designed to improve segmentation performance.
2. The perceptual features, namely, saliency, ballness, and orientation, are generated from the images without leveraging any domain knowledge using the tensor voting framework (Medioni et al., 2005).
3. The perceptual features are provided as prior information to the proposed U-shaped encoder–decoder model to improve the instance cell separation capability of the model.
4. A novel combination loss function is formulated from the focal loss and Jaccard coefficient to take care of the class imbalance problem of the dataset.

In this paper, the efficacy of the SBU-net is experimentally proved using two bright-field microscopy datasets and one fluorescence microscopy dataset. SBU-net showed significant improvement in the segmentation metrics and can lead to more research on the effective use of perceptual features for segmentation.

The remainder of this paper is structured as follows: Section 2 discusses the related works. Section 3 describes the proposed methodology in detail, followed by the experimentation and results in Section 4, discussion in Section 5 and the conclusion in Section 6.

2. Related works

The automation of microscopic image analysis involves the cell localization and segmentation step. It is the most critical and challenging step in the image analysis pipeline. A brief review of the existing image segmentation methods is conducted, and the methods are categorized into three: traditional, deep learning-based segmentation methods and prior integrated deep learning methods. For this review, we focused on the significant studies that brought architecture-level innovations for image segmentation. We have used the search strings, 'image segmentation', 'medical image segmentation', 'microscopic image segmentation' for retrieving the articles. The strengths and weaknesses of these methods concerning medical image segmentation are pointed out in Table 1 and are discussed in detail in the below paragraphs.

In traditional methods, before the advent of deep learning, the segmentation was done based on thresholding (Kong et al., 2015), clustering (Kothari et al., 2009), morphology (Gopakumar et al., 2016) and watershed (Gopakumar et al., 2018). These methods require predefined expert knowledge for segmentation. The methods fail for low-contrast images and are adversely affected by image artifacts, variation in illumination, noise, etc. These may also require seed points for cell segmentation.

In the recent past, the focus has shifted to data-driven models with advances in deep learning technology (Gopakumar and Sai Subrahmanyam, 2019). Image segmentation is now considered a pixel-level classification problem using labeled pixels. With the introduction of fully convolutional networks (FCN) by Shelhamer et al. (2017), convolutional neural networks are applied to image segmentation. Different architectures followed, namely, convolutional networks with graphical models, encoder–decoder based models, multiscale and pyramid network based models, region based convolutional neural network (RCNN) based models, deep lab family of models, recurrent neural network based models, attention based models, and so on (Minaee et al., 2022). Medical image segmentation has also progressed from these advances in image segmentation. Numerous studies have been done in the medical domain for segmentation and detection of diseases such as brain tumors from magnetic resonance imaging (MRI) images, hemorrhage in retinal images (Srinivas et al., 2022; Alam et al., 2022a,b; Sachin Saj et al., 2021; Rani et al., 2019). Improved results over the convolutional neural network were observed when FCN was used for brain image segmentation by Nie et al. (2016). FCN and its variants have shown notable performance in many studies (Wu et al., 2017; Zhou et al., 2017b,a). U-Net was developed by Ronneberger et al. (2015) for biomedical image segmentation. It was designed to specifically address the scarcity of training images in the medical field. Recently, it has emerged as a research hotspot for medical image segmentation, and many studies have used it as a backbone for various applications (Dong et al., 2017; Jin et al., 2019; Li et al., 2020). Based on differences in the function blocks used in the network architecture, the popular variants of U-Net are attention U-Net, Inception U-Net, Residual U-Net, Recurrent U-Net, Dense U-Net, U-Net++, and adversarial U-Net (Siddique et al., 2021). These exhibit higher performance than base U-Net, by addressing some of the existing problems of base U-Net, but compromising on the model complexity and cost.

Despite these advances in deep learning models for segmentation, separating the individual instances of segmented objects is an important challenge for biomedical data. In deep contour aware network (DCAN) (Chen et al., 2016), the contours of objects are predicted to separate the objects. The contour pixels may not be correctly predicted, leading to incomplete boundaries. Mask R-CNN (He et al., 2017) refines the bounding box to predict the instances after object detection. These may fail when the bounding boxes overlap. Another study known as Stardist (Schmidt et al., 2018) uses star-shaped polygon representation of objects, for instance segmentation. It requires a star-convex polygon representation for a cell and uses the distance to the boundary of each object from each pixel. It is computationally complex and memory intensive. Even though it is suitable for round shaped objects, it cannot predict objects with holes. These methods require a large number of annotated training images for good segmentation performance. In addition, a generalized approach that takes care of heterogeneous shapes of cells, cell clusters, intracellular variability, and image artifacts is still not present.

Table 1
Analysis of existing segmentation models in the literature concerning medical image segmentation.

Studies	Strengths	Weaknesses
Traditional image segmentation methods (Gopakumar et al., 2018, 2016; Kong et al., 2015)	- A large number of annotated data is not required for training - Works for images with high contrast and uniform intensity distribution	- Does not work well with low contrast medical images with non-uniform intensity distribution and artifacts - May require seed points for cell segmentation
Deep learning based segmentation models (Ronneberger et al., 2015; Siddique et al., 2021; Schmidt et al., 2018)	- Segments cells from low contrast and non-uniform intensity images with appropriate training data - Domain knowledge is not required for segmentation	- Require a large number of annotated training images - A general model that takes care of heterogeneous shapes of cells, cell clusters, intracellular variability, and image artifacts is still not present
Prior integrated deep segmentation models (Boutillon et al., 2020; Bohlender et al., 2021; Oktay et al., 2018)	- Works well with medical images having noise, artifacts and with varied anatomical shapes of cells	- May require mathematical modeling for extracting prior information - Integrating the prior information is a challenge - Model complexity may increase

Recently, there has been a paradigm shift towards incorporating prior knowledge in deep learning models. Higher segmentation accuracy is achieved by incorporating anatomical constraints as prior information (Boutillon et al., 2020; Oktay et al., 2018). The challenge lies in deciding how to include the prior information rather than modeling the prior. These are extremely useful, especially in medical images with low contrast, noise, and other artifacts. The prior information can be conditional random fields, shape models, active contour models, or topology-based models. These are applied during pre/post-processing or throughout the training process (Bohlender et al., 2021). A drawback is the requirement of mathematical modeling for the prior, which may increase the computational complexity.

It was observed that the studies on the segmentation of microscopy images are fewer than in other medical imaging modalities. U-net is applied to cell segmentation in microscopic images by Al-Kofahi et al. (2018). Tran et al. used the SegNet architecture for blood cell segmentation (Tran et al., 2018). Another work was reported on phase contrast microscopy images using U-net (Ayanzadeh et al., 2019). A combination of U-net and watershed is used for dense cell region segmentation in Lux and Matula (2019). The significant challenges in cell segmentation in microscopic images are separating overlapping/touching cells and demarcating cell boundaries. Studies have been done on predicting boundaries or separating instances in cell images (Oda et al., 2018; Yi et al., 2019). Most studies focus on fluorescence-based microscopy, in which staining makes cell segmentation easier (Sundara and Aarthi, 2019). For bright-field images, a graph partitioning approach was used to segment muscle cells (Mohiuddin and Wan, 2018). Ren et al. developed Cellbow, which uses a multilayer training set strategy with deep neural networks that work on bright-field images also Ren et al. (2020). Residual U-Net has been applied to bright-field images of retinal pigment epithelial stem cells by Patel et al. (2019). U-net has been used as a base network for semantic segmentation of yeast cells (Salem et al., 2021) from bright-field images. As is evident from the brief literature review, automation of bright-field cell segmentation is seldom studied, mainly due to the scarcity of annotated data.

In summary, the literature shows that more advanced architectures need to be developed to overcome the limitations concerning the availability of annotated training data and the inconsistencies of applying post-processing operations in improving segmentation outputs across varied images. Hence, this study proposes a novel framework for segmenting bright-field images. The perceptual features of an image can distinguish the salient regions of an image. These features can be derived from the data without domain knowledge and are incorporated as prior information into a carefully designed U-shaped encoder–decoder based segmentation model. The effect of using image based perceptual features to improve the performance of existing segmentation models is also evaluated in the study. The model has been validated with publicly available bright-field microscopy datasets and has produced the highest segmentation accuracy compared to other state-of-the-art models (Table 9).

3. Methodology

Accurate segmentation of cellular structures is essential for cell counting and determining cell morphology, where the segmented cells are used for classification and cell tracking. Automated segmentation of cells from low contrast microscopic images becomes problematic when the cells occur as dense regions with overlapping or touching cells. To enhance the performance of deep neural networks and to address these issues, some methods have used domain knowledge as prior information (Tofighi et al., 2019). However, they require manual effort for the extraction of domain knowledge.

To address these problems, a novel feature fusion approach that combines the perceptual features of the image with a segmentation framework is proposed. These features signify salient local features like curves, surfaces, junctions, and orientations of the cells in the microscopic image. Saliency, ballness, and orientation features are derived from the structure tensor of the image using a tensor voting mechanism (Medioni et al., 2005), which corresponds to the salient surfaces, curves, or junctions and orientation of the cells present in the image. The tensor voting framework, as established in the literature, has the power to complete perceptual curves. In a cell segmentation setting, this plays a significant role in segmenting cells when the cell overlaps. Tensor voting has been applied to problems such as membrane detection in electron tomography images (Martinez-Sanchez et al., 2014), vessel centerline detection in cerebral angiograms (Ding et al., 2016), and adherens detection in microscopic images (Loss et al., 2011) in the medical domain. To the best of our knowledge, no studies have been done on using the tensor voting mechanism for cell segmentation problems. The feasibility of using tensor-voted features with convolutional neural networks is not explored yet. In the proposed model, the computed features are appended to the input image in the feature fusion module, and segmentation predictions are generated using an optimized U-shaped convolutional network in an end-to-end manner.

Fig. 1 shows the proposed segmentation framework. This segmentation model comprises the following stages: computation of tensors, voting and extraction of tensors, feature fusion and preprocessing, and segmentation using a U-shaped encoder–decoder. These are discussed in detail in the following sections.

3.1. Computation of tensors

Tensor voting(TV) (Medioni et al., 2005) is a computational framework that addresses perceptual organization problems based on human perception principles. The origins of TV can be traced back to early computer vision problems in which the available input is usually sparse and noisy, making it challenging to extract meaningful information and structures. TV finds local feature descriptions by spreading shape-related input information across a neighborhood. It refines the data and emphasizes distinctive local features.

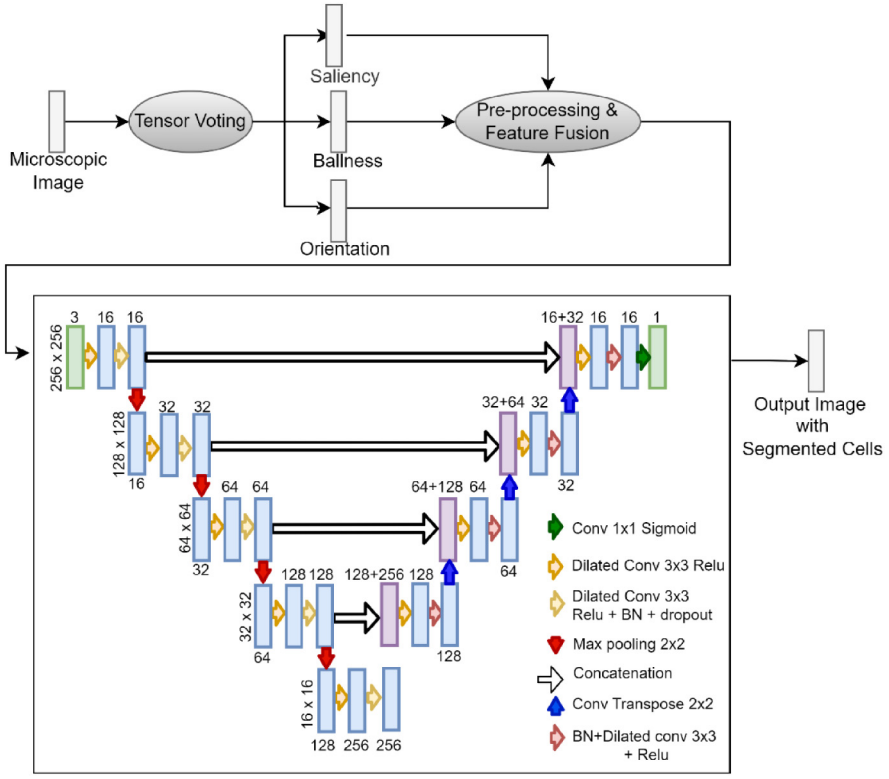


Fig. 1. SBU-Net: The input image is subjected to tensor voting operation for the computation of tensors, and then the image and tensor feature maps are pre-processed and fused. These are given as input to the U-shaped encoder–decoder model for segmentation.

Each pixel in the image may correspond to a perceptual structure like curves, surfaces, points, or junctions. The perceptual information from the small neighborhood of each pixel value can be encoded as an ellipse, whose shape and size correspond to the type of structure to which the pixel belongs. We know that any ellipse can be entirely characterized by a second-order symmetric matrix which can be further decomposed using its Eigen values and Eigen vectors as in Eq. (1).

$$\mathrm{T} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \lambda_1 \hat{e}_1 \hat{e}_1^T + \lambda_2 \hat{e}_2 \hat{e}_2^T \tag{1}$$

where λ_1 and λ_2 ($\lambda_1 \geq \lambda_2 \geq 0$) are nonnegative Eigenvalues and e_1 and e_2 are orthonormal Eigenvectors. These tensors are visualized in the elliptical form in 2D. The ellipse's shape and size are determined by the Eigenvalues, while the Eigenvectors specify its orientation. The tensor can be divided into two components, defined by Eq. (2).

$$\mathrm{T} = (\lambda_1 - \lambda_2)\hat{e}_1 \hat{e}_1^T + \lambda_2(\hat{e}_1 \hat{e}_1^T + \hat{e}_2 \hat{e}_2^T) \tag{2}$$

The saliency of the tensor is given by $\lambda_1 - \lambda_2$. It gives the measure of certainty in its encoded direction. The 'junctionness' of the tensor is given by λ_2, which measures the uncertainty in the encoded direction. Thus, without domain knowledge, the local features of the pixels are represented in the tensor form. The 'stick', 'ball' and 'orientation' tensors that encode the curve, junction and orientation information at pixel x are given by Eq. (3) where e_x is the Cartesian basis vector in x-direction. These are initially computed from Eigenvalues and Eigenvectors of the structure tensor of the input image in our model.

$$\begin{aligned} \mathrm{S}(\mathbf{x}) &= \mathrm{S}(\mathrm{T}(\mathbf{x})) = (\lambda_1 - \lambda_2)\hat{e}_1 \hat{e}_1^T \\ \mathrm{B}(\mathbf{x}) &= \mathrm{B}(\mathrm{T}(\mathbf{x})) = \lambda_2(\hat{e}_1 \hat{e}_1^T + \hat{e}_2 \hat{e}_2^T) \\ \mathrm{O}(\mathbf{x}) &= \mathrm{O}(\mathrm{T}(\mathbf{x})) = arccos(e_1 \cdot e_x) \end{aligned} \tag{3}$$

3.2. Voting and extraction of tensors

The tensors are propagated in a neighborhood using Gestalt psychology (Wertheimer, 1938) concepts of proximity, good continuation, and similarity. The input pixels stored as tensors interact locally with their neighbors by casting votes. Votes are cast such that mutually compatible tokens reinforce each other while the value of not mutually compatible tokens is ignored. The tensor vote acquired at a single position **p** is then produced by integrating the votes collected from all nearby points defined by **N(p)**. The saliency (stick), orientation, and ballness maps may be found by decomposing the resulting tensor field according to Eq. (3). The Eigenvalues describe the related saliencies, whereas the Eigenvectors define the direction of the specific features.

The input tensor field is computed, which consists of stick tensors and ball tensors at each position. The voting positions within the distance threshold cast their votes at the receiver position. It is done by aligning the voter with the receiver. The magnitude of the vote depends on the confidence that the voter and receiver belong to the same perceptual structure. In the original implementation of tensor voting, the voting field needs to be precomputed and stored. As this is not memory efficient, we adopted the voting mechanism proposed by Franken et al. (2006). This method modifies the voting mechanism to convolution using steerable filters. Theoretically, a steerable filter may be oriented in any direction by linearly combining a finite number of predefined rotations of basis filters. The output saliency is calculated using Eq. (4). The $k_m(\alpha)$ in Eq. (4) are the linear coefficients that depend on the rotation angle α, m is the index of basis filters, $V_m(x)$ are the basis filters, adapted as in Martinez-Sanchez et al. (2014). $k_m(\alpha)$ and $V_m(x)$ are defined as $k_m(\alpha) = e^{-2i(m-1)\alpha}$ and $V_m(\mathbf{x}) = \gamma_m e^{-\frac{x^2+y^2}{2\sigma_v^2}\left(\frac{x+iy}{\sqrt{x^2+y^2}}\right)^{2m}}$ for $\mathbf{x} = (x, y) \neq (0, 0)$, where $\gamma_m = \{1, 4, 6, 4, 1\}$ for $m = 0 \ldots 4$ and σ_v is the effective neighborhood size, respectively.

$$S_{output}(\mathbf{x}) = \left| \sum_{m=0}^{4} k_m(\alpha_{in}(\mathbf{x}))(S_{input}(\mathbf{x}) * V_m(\mathbf{x})) \right| \tag{4}$$

The output saliency, orientation, and ballness are calculated using this context-enhanced output saliency field.

3.3. Feature fusion and preprocessing

In the feature fusion module, the computed features in Section 3.2, namely, saliency, ballness, and orientation, are appended with the corresponding input image. The feature maps are of the same size as the input image. These are combined with the input image to create multi-channel images such that channel 0 corresponds to the input image and the remaining channels 1, 2, and 3 correspond to the saliency, ballness, or orientation feature maps. The multichannel feature rich images are fed as the input to the encoder–decoder for generating segmentation predictions. The optimal combination of saliency, ballness, and orientation feature maps for each dataset are selected based on extensive ablation studies as discussed in Section 4.4.2. Before feature fusion, the input images and computed feature maps are subjected to min–max normalization in the preprocessing module to normalize the pixel value range of the images and feature maps.

3.4. Segmentation using U-shaped encoder–decoder

Encoder–decoder architecture is selected as the backbone network for segmenting cells in the proposed model. Encoder–decoder based segmentation models are widely used for semantic segmentation due to their simple architecture. In semantic segmentation, each pixel in the image is assigned a label corresponding to its class. Since the introduction of U-net by Ronneberger et al. (2015), many variants of U-shaped encoder–decoder models have come up, as discussed in 2. The applicability of the U-shaped encoder–decoder model for semantic segmentation of small datasets of bright-field microscopic images is established in Section 4. Hence, we have chosen the U-shaped encoder–decoder model as our segmentation backbone.

The basic building blocks of U-shaped networks are the encoder and decoder. Multichannel images generated in Section 3.3 are given as input to the first encoder. The deep learning components, like convolutional layers and pooling layers can be visualized as a U-shaped pattern inside the encoder and decoder. The encoder compresses data into a lower-dimensional representation. It is also called the contraction path. The decoder then converts this data to the original image dimension. It is also called the expansion path. A single encoder block comprises two convolutional layers followed by a max pooling layer. In the decoder block, the up-convoluted feature maps are concatenated with feature maps from the encoder and are given to two convolutional layers. Skip connections are vital to the U-shaped design since they allow information to pass directly from each encoder block to the corresponding decoder block. It allows us to concatenate intermediate low-level feature map information from the encoder side with the feature maps in the decoder. The padding is kept to the 'same' to maintain the exact size of the output image as the input image. In the final layer, a convolutional layer with a filter size of 1×1 classifies each pixel into its corresponding class and generates the segmentation map.

In the proposed architecture, another addition is the technique to increase the receptive fields to operate in a global context of the cells. However, increasing the receptive field by increasing the filter size has a detrimental effect because it increases the number of parameters and the chance of overfitting. Instead, we have proposed to use dilated convolution (Wang et al., 2022; Yu and Koltun, 2015) to operate in a larger context of cells for more features that favor segmentation. The effect of dilated convolution on cell segmentation is detailed in Section 4.4.2. We have observed that this change in the architecture has improved the segmentation metrics over standard convolutions, as discussed in Section 4.4.2.

Another challenge in building an architecture for cell segmentation is the class imbalance between foreground and background pixels. The cells occupy only a tiny percentage of the total area in the microscopic images. The cross-entropy loss used in classical U-net for pixel-based segmentation will not properly converge in such cases. So, we have proposed a combination loss function (Eq. (7)) involving the Focal loss (Lin et al., 2017) and the Jaccard loss (Polak et al., 2009) to tackle this problem, and its effectiveness is discussed in Section 4.5. The Focal loss given in Eq. (5) is a variation of the cross-entropy loss function that emphasizes hard negatives while down-weighting easy ones. The Jaccard loss given in Eq. (6) is a geometric loss derived from the IoU metric and emphasizes the overlap between semantic predictions and ground truth masks. In Eqs. (5), (6) and (7), FL is the focal loss term, y_{true} is the ground truth, y_{pred} is the prediction, α is the weighting factor, and γ is the focusing parameter.

$$\begin{aligned} FL(y_{true}, y_{pred}) = &-y_{true}\alpha(1 - y_{pred})^{\gamma} \log(y_{pred}) \\ &-(1 - y_{true})\alpha y_{pred}^{\gamma} \log(1 - y_{pred}) \end{aligned} \tag{5}$$

$$J(y_{true}, y_{pred}) = 1 - \frac{y_{true} \cap y_{pred}}{y_{true} \cup y_{pred}} \tag{6}$$

$$Loss(y_{true}, y_{pred}) = FL(y_{true}, y_{pred}) + J(y_{true}, y_{pred}) \tag{7}$$

The proposed architecture is shown in Fig. 1. The segmentation model is an optimized version of the encoder–decoder architecture with added perceptual features. The proposed model differs from the baseline U-net model in the number of layers, filters, filtering operations, padding technique, and loss function. Specifically, it differs from the classical model in the following aspects.

1. In the classical version, standard convolutional layers are used. In the proposed model, dilated convolutional layers are used to enlarge the receptive field without increasing the number of parameters, which enhances the cell segmentation performance.
2. Batch normalization and dropout layers are used to prevent overfitting of the model.
3. Network depth, number of filters, and padding technique are tuned and optimized to improve the segmentation performance.
4. The combination loss function addresses the class imbalance of pixels in the image.

4. Experimental results

The proposed model is evaluated to demonstrate its applicability to cell segmentation. The current study uses bright-field microscopic images to test the model for pixel-level segmentation tasks. We have also tested the architecture on a fluorescence dataset to assess the generalization capability on a different microscopy type. The model is compared against the state-of-the-art models (Zhou et al., 2019; Schlemper et al., 2019; Chaurasia and Culurciello, 2017; Ronneberger et al., 2015) for image segmentation in terms of prediction performance. Furthermore, comprehensive ablation studies are conducted to select loss function, the optimal combination of saliency, ballness, and orientation features, and the usage of dilated convolutions. The components of the proposed model are selected based on extensive ablation experiments. The instance segmentation capability of the model is also elaborated.

The dataset details, evaluation metrics, and training details are discussed in the following sections. The results of ablation experiments and the segmentation results of the proposed model are provided in the subsequent sections.

4.1. Datasets

The segmentation model is evaluated using three datasets of microscopic images. The details of the dataset are summarized in Table 2. The manually segmented ground truth segmentation masks are available for all datasets. The cells vary across the datasets in size, texture, and appearance, and a few sample images are shown in Fig. 2.

Table 2
Details of the datasets used in the study.

Sl. No.	Dataset	Type	No. of images	Image size
1	Mouse $CD4^+$ T cells (Roy and Jacquemet, 2020)	Bright-field	209	1024 × 1024
2	Pancreatic cancer cells (Follain and Jacquemet, 2020)	Bright-field	57	1400 × 1400
3	MCF10DCIS.com cells labeled with Sir-DNA (Jacquemet, 2020)	Fluorescence	72	1024 × 1024

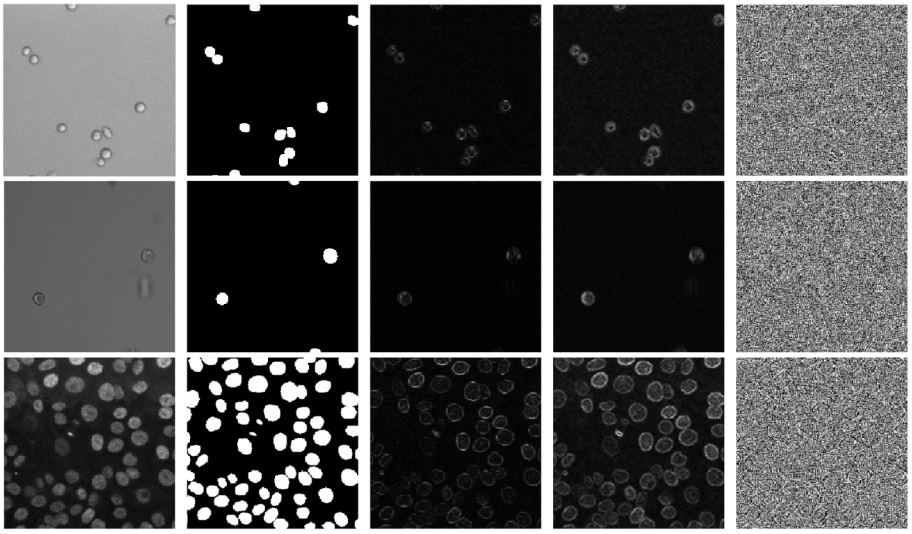


Fig. 2. Sample images, ground truth masks, and feature maps along columns (a)–(e) for datasets 1, 2, and 3. (a) Input Image (b) Ground Truth (c) Saliency Map (d) Ballness Map (e) Orientation Map.

4.2. Evaluation metrics

Segmentation is a dense prediction task. The goal is to maximize the overlap between predicted and ground truth cells. Intersection over Union (IoU) and Dice coefficient are used as the evaluation metrics to measure the overlap between predictions and ground truth. The IoU and Dice are positively correlated with values ranging from 0 to 1, with 1 indicating perfect segmentation. Compared to the Dice metric, the IoU metric gives more weight to false positives. The Dice score better represents the average segmentation performance, whereas the IoU represents the worst-case scenario.

IoU is the ratio of the number of pixels sharing predicted segmentation maps and ground truth masks to the total number of pixels in both. It is specified by Eq. (8) in terms of True Positives (TP), False Positives (FP), and False Negatives (FN). TP is the number of cell pixels correctly labeled, and FP and FN are the number of incorrectly labeled pixels for cells and background, respectively. The Dice coefficient is calculated by dividing two times the area of overlap by the total number of pixels in both images and is defined by Eq. (9). IoU and Dice scores of semantic segmentation predictions are calculated independently for each class and then averaged across all classes to get a global mean IoU(mIoU) and mean Dice(mDice) score.

$$IoU = \frac{TP}{TP + FP + FN} \tag{8}$$

$$Dice = \frac{2TP}{2TP + FP + FN} \tag{9}$$

4.3. Training details

The images in each of the datasets contain the same type of cells. Therefore, a binary segmentation is performed to classify each pixel as a cell or background from the multichannel images. The proposed model is implemented and experimented with the datasets given in Section 4.1 using Python v3.7 as the programming language. The model is implemented using Keras v2.3.1 and TensorFlow v2.1 libraries on a single GPU machine with 16 GB RAM and Nvidia Geforce GTX-1050 Ti.

CNN method is sensitive to the value of its hyperparameters. As the search space was large and different combinations of hyperparameters were possible, we did a randomized search for tuning the hyperparameters in our model. After tuning, the depth of encoder–decoder blocks is kept at three, and the number of filters is selected as (12, 32, 64, 128). The model is optimized using Adam optimizer (Kingma and Ba, 2015). Training has been done with a batch size of 8 for 100 epochs at a constant learning rate of 0.001. Batch normalization and dropout layers are used to prevent overfitting the model. The combination loss as given in Section 3.4 is used. The output segmentation maps are evaluated with the metrics given in Section 4.2.

Images are divided into equal-sized patches for training. The dataset is divided into training and validation sets in the ratios 70:30, 70:30, and 80:20 by selecting samples at random for datasets 1, 2, and 3, respectively. For the comparison with state-of-the-art models, the models are trained using the k-fold cross-validation technique. The training and validation sets are divided into patches of size 256 x 256.

4.4. Ablation experiments

A novel combination of the loss function, perceptual features, and dilated convolutions is used in the proposed model. In this section, the significance of each selected component is experimentally evaluated. The proposed combination loss function is compared against the commonly used cross-entropy loss, Focal loss, and Jaccard loss in Section 4.4.1. In Section 4.4.2, the effects of saliency, ballness, orientation features, and their combinations are investigated to find the optimal features for segmentation. The perceptual features are also compared against the traditional image features. In addition, the significance of dilated convolutions is also evaluated in Section 4.4.3.

4.4.1. Optimal selection of loss function

The loss function plays a significant role in converging a deep learning model. The segmentation of cells from microscopy images is

Table 3
Comparison of segmentation performance using different loss functions.

Loss	Dataset 1		Dataset 2		Dataset 3	
	IoU	Dice	IoU	Dice	IoU	Dice
Binary cross entropy	0.843	0.915	0.773	0.871	0.862	0.926
Focal loss	0.483	0.656	0.392	0.560	0.608	0.747
Jaccard loss	0.851	0.920	0.817	0.898	0.900	0.947
Proposed combination loss	**0.918**	**0.957**	**0.822**	**0.901**	**0.904**	**0.949**

an imbalanced class problem. Hence, as discussed in Section 3.4, a combination of Focal and Jaccard loss is proposed for the network. The model's performance using the loss function is compared against binary cross-entropy, Focal loss, and Jaccard loss. The results are summarized in Table 3. It can be concluded that using the combination loss converges the model significantly better than using other loss functions. The Jaccard loss ensures geometric overlap between predictions and masks, and Focal loss takes care of the class imbalance problem simultaneously.

4.4.2. Optimal combination of local features

The significance of the computed local features is examined extensively. The saliency feature map enhances the salient curves or surfaces in the image, the ballness feature map denotes the isolated points or junctions, and the orientation feature map denotes the orientation of structures in the image. The computed feature maps are shown in Fig. 2.

The feature maps are appended with the input image in all possible combinations to identify the most favorable feature combination for each dataset. Initially, each of the three feature maps is appended separately with the input image and evaluated. Then, two features are appended with the input image in all possible permutations and evaluated. Finally, the model is evaluated using the combination of all three feature maps with the input image (4-channel image).

The mean IoU and mean Dice scores for the experiments are listed in Tables 4 and 5. The experiments are run for 50 epochs by keeping all the hyperparameters the same. Fusion of the input image with the saliency and ballness maps gives the best performance in terms of both IoU and Dice, as can be seen in Tables 4 and 5. Orientation features are directed towards the tangent to the underlying structure, and it is observed that this feature has a detrimental effect on segmentation. At the same time, it can be seen from Tables 4 and 5 that the ballness feature positively impacts cell segmentation due to the effect of identifying junctions, thereby enhancing cell boundaries, especially when the cells overlap. Also, the addition of saliency map helps in identifying the salient curves.

The effect of perceptual features in segmentation is also compared with traditional image feature descriptors in computer vision. Histogram of gradients(HOG) (Dalal and Triggs, 2005) is an image feature descriptor widely used for object detection. The HOG features are computed for the images and appended with the input image in the same way we used for perceptual features. The results of applying HOG features for cell segmentation in three datasets are summarized in Table 6. As expected, the saliency and ballness features capture the salient features of the cells in a better way than the HOG features.

4.4.3. Effect of dilated convolutions

Dilated convolutions expand the receptive field exponentially without increasing the number of parameters or the computational requirement. The proposed model utilizes dilated convolutions to capture its advantages. Standard convolutions are replaced with dilated convolutions in the encoder and decoder. The segmentation performance with and without dilated convolutions is quantified in Table 7.

The experiments are conducted for 50 epochs by keeping the same hyperparameters for the three datasets. The proposed model with dilated convolutions achieves a mean IoU of 0.918, 0.822, 0.9038 and a mean Dice of 0.957, 0.901, 0.949 for datasets 1, 2, and 3, respectively. The mean IoU is improved by 1.55%, 0.82%, and the mean Dice is improved by 0.84%, 1.44% by using dilated convolutions for the bright-field datasets 1 and 2. The IoU and Dice also show a percentage increase of 0.055% and 0.01% for the fluorescent dataset. It indicates that dilated convolutions capture more global context than standard convolutions, especially for bright-field datasets. Hence, it can be concluded from the experiments that the application of dilated convolutions positively affects the segmentation performance of all three datasets.

4.5. Comparison with state-of-the-art methods for semantic segmentation

The proposed model is termed Saliency and Ballness driven U-shaped Network (SBU-net). This section compares SBU-net with state-of-the-art deep learning techniques in image segmentation, namely U-net (Ronneberger et al., 2015), Link-net (Chaurasia and Culurciello, 2017), Attention U-net (Schlemper et al., 2019) and U-net++ (Zhou et al., 2019). All these models are designed for semantic segmentation, of which U-net and Attention U-net deal with medical image segmentation. The networks are trained from scratch using feature-enhanced and original images for 100 epochs. The datasets are subjected to k-fold cross-validation to keep the performance comparison of the models valid when operating on limited data. It ensures that the training and validation are independent of the training set selection. The k value is set to 5, and the dataset is divided into five folds, keeping one fold for validation in each iteration. The rest of the four folds are used for training.

The experimentation is carried out to evaluate two aspects: the segmentation capability of SBU-net and the ability of the feature fusion module to enhance the segmentation quality. Sections 4.5.1 and 4.5.2 discusses the results in detail.

4.5.1. Assessment of segmentation capability of SBU-net

The segmentation performance of SBU-net is evaluated against the state-of-the-art models. The total number of parameters for the model is around 24% of U-net, 94% of U-net++, 2% of Link-net and 20% of Attention U-net (Table 8). The proposed model took 115 min against 200 min for Link-net on average across the three datasets when tested under similar settings. Table 9 summarizes the results of 5 fold cross-validation on the three datasets. SBU-net achieves comparatively greater IoU and Dice values for all three datasets.

The statistical significance of the proposed model is evaluated using the paired samples t-test (McCall and Kagan, 1994). It is a statistical test to ensure that the mean difference between two sets of observations differs from zero. The test is applied to the mDice and mIoU metrics of 5 fold cross-validation experiment. The significance of SBU-net is evaluated against each of the other models. Table 10 shows the p values of the test. The null hypothesis is taken as the distribution of metrics belonging to the same model, and there is not much difference between them. In all the cases, SBU-net achieved less p-value than the sigma, which is taken as 0.05. It proves and establishes the fact that the proposed model positively has a statistically significant effect on cell segmentation.

The sample image patches and the corresponding predictions for the models are shown in Fig. 3. It can be observed that SBU-net exhibited better segmentation when compared with that of U-net, U-net++, Attention U-net, and Link-net in all three datasets. The improvement in segmentation is attributed to the application of feature enhancement by tensor voting and the usage of dilated convolutions in the proposed u-shaped encoder–decoder model.

Since the separation of touching and overlapping cells is an open problem, the instance segmentation capability of SBU-net is examined qualitatively. We have shown a section of the input image with overlapping/touching cells in Fig. 3 to assess the instance segmentation capability of the proposed architecture. The output of SBU-net is shown in the last column. It can be seen that a thin boundary separates the touching cells favoring instance separation when compared to the results shown against other methods.

Table 4
Comparison of segmentation performance of different feature combinations using mIoU.

Dataset	Saliency(S)	Ballness(B)	Orientation(O)	S & B	B & O	S & O	S & B & O
1	0.910	0.916	0.910	**0.918**	0.915	0.912	0.914
2	0.818	0.820	0.817	**0.822**	0.819	0.819	0.817
3	0.872	0.879	0.878	**0.903**	0.892	0.888	0.892

Table 5
Comparison of segmentation performance of different feature combinations using mDice.

Dataset	Saliency(S)	Ballness(B)	Orientation(O)	S & B	B & O	S & O	S & B & O
1	0.953	0.956	0.953	**0.957**	0.956	0.954	0.955
2	0.898	0.899	0.898	**0.901**	0.899	0.899	0.898
3	0.931	0.935	0.935	**0.949**	0.942	0.940	0.942

Table 6
Comparison of perceptual features with traditional image feature descriptor(Histogram Of Gradients).

Feature	Dataset 1		Dataset 2		Dataset 3	
	IoU	Dice	IoU	Dice	IoU	Dice
Without any feature	0.800	0.888	0.815	0.895	0.849	0.916
HOG feature	0.885	0.939	0.811	0.894	0.886	0.939
Saliency and Ballness feature	**0.918**	**0.957**	**0.822**	**0.901**	**0.904**	**0.949**

Table 7
Comparison of segmentation performance on using dilated convolutions using mIoU and mDice.

Dataset	Dilated convolutions		Standard convolutions	
	IoU	Dice	IoU	Dice
1	0.918	0.957	0.904	0.949
2	0.822	0.901	0.815	0.897
3	0.9038	0.9494	0.9033	0.9491

Table 8
Comparison of number of parameters in SBU-net and other models.

SBU-net	U-net	U-net++	Link-net	Attention U-net
4,84,849	19,68,517	5,15,529	2,03,25,137	23,41,801

4.5.2. Assessment of feature fusion module

The efficacy of the feature fusion module for segmentation is evaluated by incorporating the module into the existing segmentation models. The significance of feature fusion in SBU-net is also determined by comparing it against the proposed model without the feature fusion component.

The effect of feature fusion on each architecture is assessed by adding feature-enhanced images in the state-of-the-art models. The segmentation performance of the models improved upon the addition of perceptual features. The results are reported in Table 11. It is inferred from the experiments that U-net, U-net++, Attention U-net, and Link-net achieve a higher segmentation performance on using tensor-voted images. From Tables 9 and 11, it can be observed that the IoU and Dice values for U-net, U-net++, Link-net and Attention U-net improved by 8.5%, 2.57%, 1.65%, 7.63% and 4.74%, 2.42%, 1.25%, 4.25% respectively for dataset 1. For dataset 2, the values improved by 2.63%, 0.62%, 0.62%, 0.99% and 2.04%, 0.34%, 0.34%, 0.56% respectively. Similarly, the values improved by 5.73%, 4.1%, 6.96%, 3.24% and 3.34%, 2.44%, 3.59%, 1.76% respectively for dataset 3.

The impact of feature enhancement using perceptual features is assessed by comparing the proposed encoder–decoder architecture using original input images and the feature-enhanced multi-channel images. The improvement in average Dice and average IoU across the five folds upon applying the feature enhancement is shown in Fig. 4. The difference indicates the significance of feature enhancement using perceptual features. The validation loss curves obtained when the datasets are trained with feature-enhanced images and original images for the SBU-net are given in Fig. 5. The curves indicate that the network can learn faster using feature-enhanced images and attain the minimum loss.

5. Discussions

The research gaps identified and challenged in the current research are heterogeneous shapes of cells, intracell variability, and the occurrence of cells as clusters. Segmenting cells from low-contrast bright-field images is also tricky using standard image processing operations. There is also a requirement for deep models that can work with few annotated training data. Hence, segmenting cells from microscopy images, mainly from bright-field microscopy, is an open problem in the literature. The present study solves the cell segmentation problem using perceptual features of the image. The perceptual features were applied to an optimized encoder–decoder model to address the challenges of

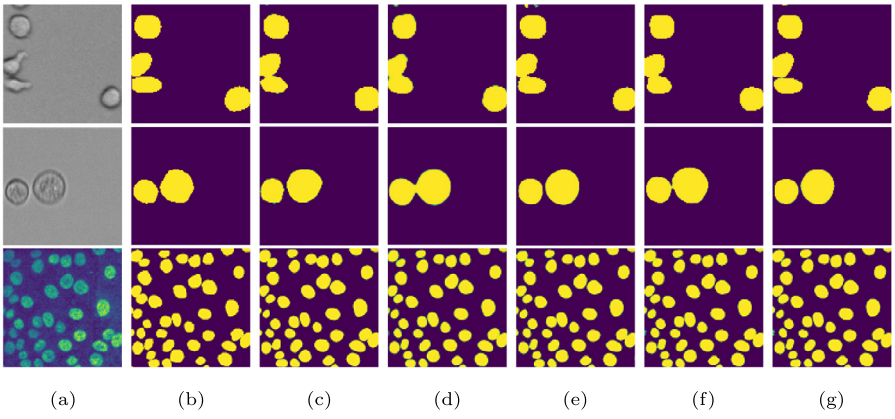


Fig. 3. Comparison of SBU-net with U-net, U-net++, Link-net, Attention U-net for datasets 1, 2 and 3. (a) Input Image (b) Ground truth mask (c) U-net (d) U-net++ (e) Link-net (f) Attention U-net (g) SBU-net.

Table 9
Comparison of SBU-net with the state-of-the-art segmentation models. The values are shown in mean ± standard deviation of mIoU and mDice scores.

Model	Dataset 1		Dataset 2		Dataset 3	
	IoU	Dice	IoU	Dice	IoU	Dice
U-net	0.729 ± 0.026	0.843 ± 0.017	0.797 ± 0.024	0.881 ± 0.020	0.820 ± 0.015	0.898 ± 0.011
U-net++	0.777 ± 0.020	0.866 ± 0.012	0.812 ± 0.007	0.895 ± 0.004	0.829 ± 0.007	0.902 ± 0.002
Link-net	0.790 ± 0.011	0.879 ± 0.008	0.812 ± 0.012	0.894 ± 0.009	0.805 ± 0.010	0.891 ± 0.007
Attention U-net	0.734 ± 0.023	0.846 ± 0.015	0.807 ± 0.015	0.892 ± 0.001	0.834 ± 0.028	0.908 ± 0.018
SBU-net	**0.804 ± 0.017**	**0.891 ± 0.011**	**0.829 ± 0.005**	**0.906 ± 0.003**	**0.892 ± 0.024**	**0.948 ± 0.023**

Table 10
p values of paired samples t-test on the 5-fold cross-validation of SBU-net and state-of-the-art segmentation models. Sigma is taken as 0.05.

Dataset	U-net		U-net++		Attention U-net		Link-net	
	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice
Dataset 1 - SBU-net	0.002	0.002	0.037	0.009	0.001	0.001	0.041	0.028
Dataset 2 - SBU-net	0.043	0.041	0.018	0.019	0.022	0.025	0.030	0.050
Dataset 3 - SBU-net	0.003	0.010	0.003	0.014	0.018	0.031	0.003	0.011

Table 11
Impact of feature fusion on segmentation models using mIoU and mDice. The values are shown in mean ± standard deviation of mIoU and mDice scores.

Model	Dataset 1		Dataset 2		Dataset 3	
	IoU	Dice	IoU	Dice	IoU	Dice
U-net	0.791 ± 0.019	0.883 ± 0.012	0.818 ± 0.012	0.899 ± 0.008	0.867 ± 0.025	0.928 ± 0.016
U-net++	0.797 ± 0.016	0.887 ± 0.010	0.817 ± 0.012	0.898 ± 0.008	0.863 ± 0.022	0.924 ± 0.015
Link-net	0.803 ± 0.014	0.890 ± 0.009	0.817 ± 0.011	0.897 ± 0.008	0.861 ± 0.027	0.923 ± 0.017
Attention U-net	0.790 ± 0.018	0.882 ± 0.011	0.815 ± 0.007	0.897 ± 0.004	0.861 ± 0.022	0.924 ± 0.014

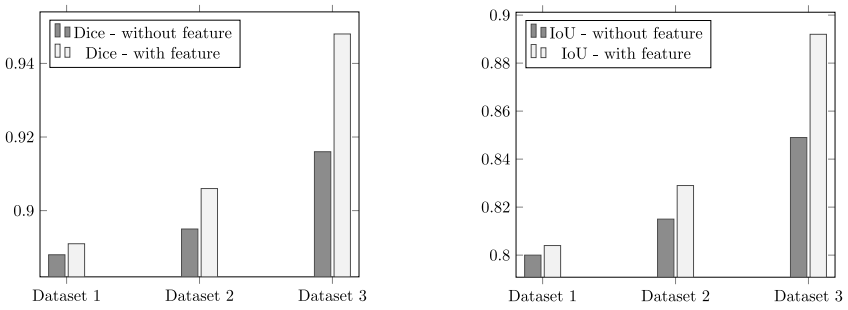


Fig. 4. Bar chart showing the improvement in (a) Dice values (b) IoU values of applying feature enhancement to the proposed model.

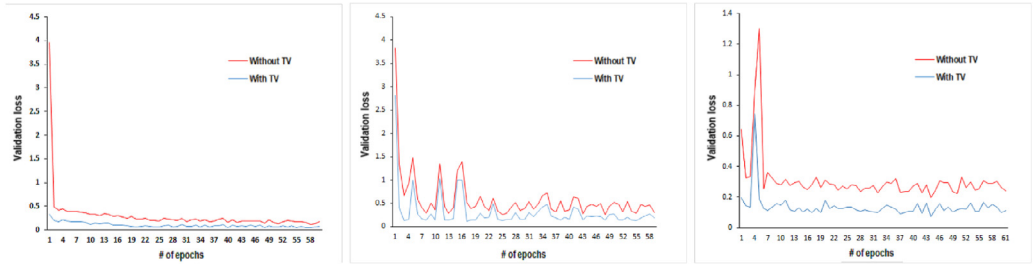


Fig. 5. Comparison of validation loss curves for SBU-net with feature enhanced images vs. without feature enhanced images for datasets 1, 2, and 3.

cell segmentation. The discussions on the obtained results are detailed below:

Segmentation performance of SBU-net: SBU-net comprises perceptual features, loss function, and dilated convolutions in place of standard convolutions. The loss function and optimal feature combination are chosen using extensive ablation experiments. The effectiveness of dilated convolutions over standard convolutions is also established experimentally. SBU-net outperforms U-net, U-net++, Link-net, and Attention U-net in terms of IOU and Dice metrics. The IoU and Dice metrics are improved by 10.3%, 4%, 8.8% and 5.7%, 2.8%, 5.6% from the baseline U-net model for the three datasets, respectively. It is attributed to the usage of saliency and ballness features and the carefully selected components of the encoder–decoder model. The feature fusion using perceptual features allowed the deep model to learn from both input images and the perceptual feature maps. It resulted in the improved segmentation quality and separation of individual instances of cells.

Dilated convolutions expanded the receptive field without increasing the complexity of the model and presented a global context of cells to the model. Batch normalization and dropout techniques made sure that the model was not overfitting. Another significant contribution is the proposed combination loss function. Jaccard loss enforced the geometrical correctness, whereas Focal loss ensured that the training was not affected by class imbalance.

Significance of feature fusion module: The significance of the feature fusion module is also experimentally evaluated for U-net, U-net++, Link-net, and Attention U-net. It led to an improvement in the segmentation metrics for all the models. The impact of the module on the proposed architecture is also assessed. Hence, the usability of the feature fusion module as plug and play component for segmentation models is established in the study. The perceptual features guide the segmentation models to learn the salient regions and enforce the ballness property of cells. It opens up an exciting research path in developing better detection, segmentation, and classification models.

6. Conclusion

The trade-off between the model complexity, accuracy, and the amount of training data is essential in developing segmentation models, especially deep learning models. It is important to integrate the strong theory and mathematically sound techniques developed by the computer vision community in the last decades to improve the performance of these deep learning models in constrained resource-limited settings. It will also add meaning to the reported high performance of the deep learning model, which often appears as a black box. We have made a significant effort in this direction for the cell segmentation problem in microscopy by showing that the perceptual features we have introduced performed better on the state-of-the-art results reported by existing models. Also, the designed architecture with the proposed loss function to tackle class imbalance issues outperformed all other models giving state-of-the-art results in semantic segmentation on the public datasets considered in this paper.

The incorporation of saliency and ballness features in the proposed SBU-net model enhanced the segmentation performance and produced superior performance on bright-field microscopic images. Perceptual feature fusion significantly improved Dice and IoU metrics for the state-of-the-art models widely used for semantic segmentation. Furthermore, it has been proved that optimizing a simple U-shaped network is better than using complex architectures to improve segmentation performance. The research findings are expected to advance research and development in automated microscopic image segmentation toward the effective use of perceptual features.

7. Future work

The model's capability to segment cells from bright-field microscopic images is established in the study. The generalization capability is also assessed using the fluorescence dataset. One research direction extending this work could be adapting the network to address multi-cell segmentation problems. Instance segmentation is particularly useful for cell counting applications and for measuring the morphological properties of cellular structures. The inclusion of perceptual features for instance segmentation could be another direction. We believe research in this aspect will open up exciting dimensions in advancing the deep learning models by utilizing the traditional computer vision knowledge bases.

CRediT authorship contribution statement

S.B. Asha: Methodology, Software, Investigation, Writing – original draft. **G. Gopakumar:** Conceptualization, Writing – review & editing, Supervision. **Gorthi R.K. Sai Subrahmanyam:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The datasets used in the study are publicly available.

References

Al-Kofahi, Y., Zaltsman, A., Graves, R., Marshall, W., Rusu, M., 2018. A deep learning-based algorithm for 2-d cell segmentation in microscopy images. BMC Bioinformatics 19 (365), http://dx.doi.org/10.1186/s12859-018-2375-z.

Alam, T.M., Shaukat, K., Khelifi, A., Khan, W.A., Raza, H.M.E., Idrees, M., Luo, S., Hameed, I.A., 2022a. Disease diagnosis system using iot empowered with fuzzy inference system. Comput. Mater. Contin. 70 (3), 5305–5319. http://dx.doi.org/10.32604/cmc.2022.020344.

Alam, T.M., Shaukat, K., Mahboob, H., Sarwar, M.U., Iqbal, F., Nasir, A., Hameed, I.A., Luo, S., 2022b. A machine learning approach for identification of malignant mesothelioma etiological factors in an imbalanced dataset. Comput. J. 65 (7), 1740–1751. http://dx.doi.org/10.1093/comjnl/bxab015.

Ali, M.A.S., Misko, O., Salumaa, S.O., Papkov, M., Palo, K., Fishman, D., Parts, L., 2021. Evaluating very deep convolutional neural networks for nucleus segmentation from brightfield cell microscopy images. SLAS DISCOVERY: Adv. Sci. Drug Discov. 26 (9), 1125–1137. http://dx.doi.org/10.1177/24725552211023214.

Ayanzadeh, A., Yağar, H.O., Özuysal, z.Y., Okvur, D.P., Töreyin, D., Önal, S., 2019. Cell segmentation of 2d phase-contrast microscopy images with deep learning method. In: 2019 Medical Technologies Congress. TIPTEKNO, pp. 1–4. http://dx.doi.org/10.1109/TIPTEKNO.2019.8894978.

Bohlender, S., Oksuz, I., Mukhopadhyay, A., 2021. A survey on shape-constraint deep learning for medical image segmentation. http://dx.doi.org/10.48550/ARXIV.2101.07721.

Boutillon, A., Borotikar, B., Burdin, V., Conze, P.H., 2020. Combining shape priors with conditional adversarial networks for improved scapula segmentation in mr images. In: 2020 IEEE 17th International Symposium on Biomedical Imaging. ISBI, pp. 1164–1167. http://dx.doi.org/10.1109/ISBI45749.2020.9098360.

Chaurasia, A., Culurciello, E., 2017. Linknet: Exploiting encoder representations for efficient semantic segmentation. In: 2017 IEEE Visual Communications and Image Processing. VCIP, pp. 1–4. http://dx.doi.org/10.1109/VCIP.2017.8305148.

Chen, H., Qi, X., Yu, L., Heng, P.A., 2016. Dcan: Deep contour-aware networks for accurate gland segmentation. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 2487–2496. http://dx.doi.org/10.1109/CVPR.2016.273.

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). pp. 886–893. http://dx.doi.org/10.1109/CVPR.2005.177.

Dimopoulos, S., Mayer, C.E., Rudolf, F., Stelling, J., 2014. Accurate cell segmentation in microscopy images using membrane patterns. Bioinformatics 30 (18), 2644–2651. http://dx.doi.org/10.1093/bioinformatics/btu302.

Ding, Y., Nicolescu, M., Farmer, D., Wang, Y., Bebis, G., Scalzo, F., 2016. Tensor voting extraction of vessel centerlines from cerebral angiograms. In: Advances in Visual Computing. Springer International Publishing, Cham, pp. 35–44. http://dx.doi.org/10.1007/978-3-319-50835-1_4.

Dong, H., Yang, G., Liu, F., Mo, Y., Guo, Y., 2017. Automatic brain tumor detection and segmentation using u-net based fully convolutional networks. In: Valdés Hernández, M., González-Castro, V. (Eds.), Medical Image Understanding and Analysis. Springer International Publishing, Cham, pp. 506–517. http://dx.doi.org/10.1007/978-3-319-60964-5_44.

Follain, G., Jacquemet, G., 2020. Combining StarDist and TrackMate example 3 - flow chamber dataset. http://dx.doi.org/10.5281/zenodo.4034939.

Franken, E., van Almsick, M., Rongen, P., Florack, L., ter Haar Romeny, B., 2006. An efficient method for tensor voting using steerable filters. In: Leonardis, A., Bischof, H., Pinz, A. (Eds.), Computer Vision – ECCV 2006. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 228–240. http://dx.doi.org/10.1007/11744085_18.

Gopakumar, G., Jagannadh, V.K., Gorthi, S.S., Subrahmanyam, G.R., 2016. Framework for morphometric classification of cells in imaging flow cytometry. J. Microsc. 261 (3), 307–319. http://dx.doi.org/10.1111/jmi.12335.

Gopakumar, G., Sai Subrahmanyam, G.R.K., 2019. Deep learning applications to cytopathology: A study on the detection of malaria and on the classification of leukaemia cell-lines. In: Balas, V.E., Roy, S.S., Sharma, D., Samui, P. (Eds.), Handbook of Deep Learning Applications. Springer International Publishing, Cham, pp. 219–257. http://dx.doi.org/10.1007/978-3-030-11479-4_11.

Gopakumar, G.P., Swetha, M., Sai Siva, G., Sai Subrahmanyam, G.R.K., 2018. Convolutional neural network-based malaria diagnosis from focus stack of blood smear images acquired using custom-built slide scanner. J. Biophotonics 11 (3), e201700003. http://dx.doi.org/10.1002/jbio.201700003.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn. In: 2017 IEEE International Conference on Computer Vision. ICCV, pp. 2980–2988. http://dx.doi.org/10.1109/ICCV.2017.322.

Jacquemet, G., 2020. Combining StarDist and TrackMate example 1 - breast cancer cell dataset. http://dx.doi.org/10.5281/zenodo.4034976.

Jin, Q., Meng, Z., Pham, T.D., Chen, Q., Wei, L., Su, R., 2019. Dunet: A deformable network for retinal vessel segmentation. Knowl.-Based Syst. 178, 149–162. http://dx.doi.org/10.1016/j.knosys.2019.04.025.

Kalmady, K.S., Kamath, A.S., Gopakumar, G., Subrahmanyam, G.R.K.S., Gorthi, S.S., 2017. Improved transfer learning through shallow network embedding for classification of leukemia cells. In: 2017 Ninth International Conference on Advances in Pattern Recognition. ICAPR, pp. 1–6. http://dx.doi.org/10.1109/ICAPR.2017.8593186.

Kingma, D.P., Ba, J., 2015. Adam: A method for stochastic optimization. doi:arXiv:1412.6980.

Kong, J., Wang, F., Teodoro, G., Liang, Y., Zhu, Y., Tucker-Burden, C., Brat, D.J., 2015. Automated cell segmentation with 3d fluorescence microscopy images. In: 2015 IEEE 12th International Symposium on Biomedical Imaging. ISBI, pp. 1212–1215. http://dx.doi.org/10.1109/ISBI.2015.7164091.

Kothari, S., Chaudry, Q., Wang, M.D., 2009. Automated cell counting and cluster segmentation using concavity detection and ellipse fitting techniques. In: 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. pp. 795–798. http://dx.doi.org/10.1109/ISBI.2009.5193169.

Li, S., Tso, G.K., He, K., 2020. Bottleneck feature supervised u-net for pixel-wise liver and tumor segmentation. Expert Syst. Appl. 145, 113131. http://dx.doi.org/10.1016/j.eswa.2019.113131.

Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal loss for dense object detection. In: 2017 IEEE International Conference on Computer Vision. ICCV, pp. 2999–3007. http://dx.doi.org/10.1109/ICCV.2017.324.

Loss, L.A., Bebis, G., Parvin, B., 2011. Iterative tensor voting for perceptual grouping of ill-defined curvilinear structures. IEEE Trans. Med. Imaging 30 (8), 1503–1513. http://dx.doi.org/10.1109/TMI.2011.2129526.

Lux, F., Matula, P., 2019. Dic image segmentation of dense cell populations by combining deep learning and watershed. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). pp. 236–239. http://dx.doi.org/10.1109/ISBI.2019.8759594.

Martinez-Sanchez, A., Garcia, I., Asano, S., Lucic, V., Fernandez, J.J., 2014. Robust membrane detection based on tensor voting for electron tomography. J. Struct. Biol. 186 (1), 49–61. http://dx.doi.org/10.1016/j.jsb.2014.02.015.

McCall, R., Kagan, J., 1994. Fundamental Statistics for Behavioral Sciences. Harcourt Brace College Publishers, San Diego.

Medioni, G., Mordohai, P., Nicolescu, M., 2005. The tensor voting framework. In: Handbook of Geometric Computing: Applications in Pattern Recognition, Computer Vision, Neuralcomputing, and Robotics. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 535–568. http://dx.doi.org/10.1007/3-540-28247-5_16.

Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., Terzopoulos, D., 2022. Image segmentation using deep learning: A survey. IEEE Trans. Pattern Anal. Mach. Intell. 44 (7), 3523–3542. http://dx.doi.org/10.1109/TPAMI.2021.3059968.

Moen, E., Bannon, D., Kudo, T., Graf, W., Covert, M., Van Valen, D., 2019. Deep learning for cellular image analysis. Nature Methods 16 (12), 1233–1246. http://dx.doi.org/10.1038/s41592-019-0403-1.

Mohiuddin, K., Wan, J.W., 2018. Automated segmentation of cellular images using an effective region force. J. Med. Imaging 5 (4), 044002. http://dx.doi.org/10.1117/1.JMI.5.4.044002.

Mualla, F., Aubreville, M., Maier, A., 2018. Microscopy. Springer International Publishing, Cham, http://dx.doi.org/10.1007/978-3-319-96520-8_5.

Nie, D., Wang, L., Gao, Y., Shen, D., 2016. Fully convolutional networks for multi-modality isointense infant brain image segmentation. In: 2016 IEEE 13th International Symposium on Biomedical Imaging. pp. 1342–1345. http://dx.doi.org/10.1109/ISBI.2016.7493515.

Oda, H., Roth, H.R., Chiba, K., Sokolić, J., Kitasaka, T., Oda, M., Hinoki, A., Uchida, H., Schnabel, J.A., Mori, K., 2018. Besnet: Boundary-enhanced segmentation of cells in histopathological images. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. Springer International Publishing, Cham, pp. 228–236. http://dx.doi.org/10.1007/978-3-030-00934-2_26.

Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M., Bai, W., Caballero, J., Cook, S.A., de Marvao, A., Dawes, T., O'Regan, D.P., Kainz, B., Glocker, B., Rueckert, D., 2018. Anatomically constrained neural networks (acnns): Application to cardiac image enhancement and segmentation. IEEE Trans. Med. Imaging 37 (2), 384–395. http://dx.doi.org/10.1109/TMI.2017.2743464.

Patel, G., Tekchandani, H., Verma, S., 2019. Cellular segmentation of bright-field absorbance images using residual u-net. In: 2019 International Conference on Advances in Computing, Communication and Control (ICAC3). pp. 1–5. http://dx.doi.org/10.1109/ICAC347590.2019.9036737.

Polak, M., Zhang, H., Pi, M., 2009. An evaluation metric for image segmentation of multiple objects. Image Vis. Comput. 27 (8), 1223–1227. http://dx.doi.org/10.1016/j.imavis.2008.09.008.

Rani, N.S., Bipin, N.B.J., Yadhu, C.R., 2019. Hemorrhage segmentation and detection in retinal images using object detection techniques and machine learning perspectives. In: 2019 Global Conference for Advancement in Technology. GCAT, pp. 1–5. http://dx.doi.org/10.1109/GCAT47503.2019.8978422.

Ren, H., Zhao, M., Liu, B.e.a., 2020. Cellbow: a robust customizable cell segmentation program. Quant. Biol. 8, 245–255. http://dx.doi.org/10.1117/1.JMI.5.4.044002.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Springer International Publishing, Cham, pp. 234–241. http://dx.doi.org/10.1007/978-3-319-24574-4_28.

Roy, N.H., Jacquemet, G., 2020. Combining StarDist and TrackMate example 2 - t cell dataset. http://dx.doi.org/10.5281/zenodo.4034929.

Sachin Saj, T.K., Sowmya, V., Soman, K.P., 2021. Performance analysis of segmentor adversarial network (segan) on bio-medical images for image segmentation. In: Komanapalli, V.L.N., Sivakumaran, N., Hampannavar, S. (Eds.), Advances in Automation, Signal Processing, Instrumentation, and Control. Springer Nature Singapore, Singapore, pp. 751–758. http://dx.doi.org/10.1007/978-981-15-8221-9_69.

Salem, D., Li, Y., Xi, P., Phenix, H., Cuperlovic-Culf, M., Kæn, M., 2021. Yeastnet: Deep-learning-enabled accurate segmentation of budding yeast cells in bright-field microscopy. Appl. Sci. 11 (6), http://dx.doi.org/10.3390/app11062692.

Schlemper, J., Oktay, O., Schaap, M., Heinrich, M., Kainz, B., Glocker, B., Rueckert, D., 2019. Attention gated networks: Learning to leverage salient regions in medical images. Med. Image Anal. 53, 197–207. http://dx.doi.org/10.1016/j.media.2019.01.012.

Schmidt, U., Weigert, M., Broaddus, C., Myers, G., 2018. Cell detection with star-convex polygons. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. Springer International Publishing, Cham, pp. 265–273. http://dx.doi.org/10.1007/978-3-030-00934-2_30.

Shaukat, K., Iqbal, F., Alam, T.M., Aujla, G.K., Devnath, L., Khan, A.G., Iqbal, R., Shahzadi, I., Rubab, A., 2020. The impact of artificial intelligence and robotics on the future employment opportunities. Trends Comput. Sci. Inf. Technol. 5 (1), 050–054. http://dx.doi.org/10.17352/tcsit.000022.

Shelhamer, E., Long, J., Darrell, T., 2017. Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39 (4), 640–651. http://dx.doi.org/10.1109/TPAMI.2016.2572683.

Siddique, N., Paheding, S., Elkin, C.P., Devabhaktuni, V., 2021. U-net and its variants for medical image segmentation: A review of theory and applications. IEEE Access 9, 82031–82057. http://dx.doi.org/10.1109/ACCESS.2021.3086020.

Simon, A., Vinayakumar, R., Sowmya, V., Soman, K.P., 2019. Shallow cnn with lstm layer for tuberculosis detection in microscopic image. Int. J. Recent Technol. Eng. 7, 56–60, URL: https://www.ijrte.org/wp-content/uploads/papers/v7i6s3/F1012376S19.pdf.

Song, J., Xiao, L., Lian, Z., 2018. Contour-seed pairs learning-based framework for simultaneously detecting and segmenting various overlapping cells/nuclei in microscopy images. IEEE Trans. Image Process. 27 (12), 5759–5774. http://dx.doi.org/10.1109/TIP.2018.2857001.

Srinivas, C., KS, N.P., Zakariah, M., Alothaibi, Y.A., Shaukat, K., Partibane, B., Awal, H., 2022. Deep transfer learning approaches in performance analysis of brain tumor classification using mri images. J. Healthc. Eng. 2022 (3264367), http://dx.doi.org/10.1155/2022/3264367.

Sundara, S.M., Aarthi, R., 2019. Segmentation and evaluation of white blood cells using segmentation algorithms. In: 2019 3rd International Conference on Trends in Electronics and Informatics. ICOEI, pp. 1143–1146. http://dx.doi.org/10.1109/ICOEI.2019.8862724.

Thanmayi, A.V.L., Reddy, S.D., Kochuvila, S., 2021. Detection of leukemia using k-means clustering and machine learning. In: Kumar, N., Vinodhini, M., Venkatesha Prasad, R.R. (Eds.), Ubiquitous Communications and Network Computing. Springer International Publishing, Cham, pp. 198–209. http://dx.doi.org/10.1007/978-3-030-79276-3_15.

Tofighi, M., Guo, T., Vanamala, J.K.P., Monga, V., 2019. Prior information guided regularized deep learning for cell nucleus detection. IEEE Trans. Med. Imaging 38 (9), 2047–2058. http://dx.doi.org/10.1109/TMI.2019.2895318.

Tran, T., Kwon, O.H., Kwon, K.R., Lee, S.H., Kang, K.W., 2018. Blood cell images segmentation using deep learning semantic segmentation. In: 2018 IEEE International Conference on Electronics and Communication Engineering. ICECE, pp. 13–16. http://dx.doi.org/10.1109/ICECOME.2018.8644754.

Wang, S., Singh, V.K., Cheah, E., Wang, X., Li, Q., Chou, S.H., Lehman, C.D., Kumar, V., Samir, A.E., 2022. Stacked dilated convolutions and asymmetric architecture for u-net-based medical image segmentation. Comput. Biol. Med. 148, 105891. http://dx.doi.org/10.1016/j.compbiomed.2022.105891.

Wertheimer, M., 1938. Laws of Organization in Perceptual Forms. Kegan Paul, Trench, Trubner & Company, http://dx.doi.org/10.1037/11496-005.

Wu, L., Xin, Y., Li, S., Wang, T., Heng, P.A., Ni, D., 2017. Cascaded fully convolutional networks for automatic prenatal ultrasound image segmentation. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). pp. 663–666. http://dx.doi.org/10.1109/ISBI.2017.7950607.

Xing, F., Xie, Y., Su, H., Liu, F., Yang, L., 2018. Deep learning in microscopy image analysis: A survey. IEEE Trans. Neural Netw. Learn. Syst. 29 (10), 4550–4568. http://dx.doi.org/10.1109/TNNLS.2017.2766168.

Yi, J., Wu, P., Jiang, M., Huang, Q., Hoeppner, D.J., Metaxas, D.N., 2019. Attentive neural cell instance segmentation. Med. Image Anal. 55, 228–240. http://dx.doi.org/10.1016/j.media.2019.05.004.

Yu, F., Koltun, V., 2015. Multi-scale context aggregation by dilated convolutions. doi:arXiv:1511.07122.

Zhou, X., Shen, M., Riga, C.V., Yang, G., Lee, S., 2017a. Focal FCN: towards small object segmentation with limited training data. doi:arXiv:1711.01506.

Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J., 2019. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE Trans. Med. Imaging 39 (6), 1856–1867. http://dx.doi.org/10.1109/TMI.2019.2959609.

Zhou, X., Takayama, R., Wang, S., Hara, T., Fujita, H., 2017b. Deep learning of the sectional appearances of 3d ct images for anatomical structure segmentation based on an fcn voting method. Med. Phys. 44 (10), 5221–5233. http://dx.doi.org/10.1002/mp.12480.