

## Article

# Cervical Cell Segmentation Method Based on Global Dependency and Local Attention

Gang Li <sup>1</sup>, Chengjie Sun <sup>1,\*</sup>, Chuanyun Xu <sup>2</sup>, Yu Zheng <sup>1</sup> and Keya Wang <sup>1</sup>

<sup>1</sup> School of Artificial Intelligence, Chongqing University of Technology, Chongqing 400054, China; ligang@cqut.edu.cn (G.L.); langdaoliu@2019.cqut.edu.cn (Y.Z.); 709207454@stu.cqut.edu.cn (K.W.)

<sup>2</sup> College of Computer and Information Science, Chongqing Normal University, Chongqing 401331, China; xcy@cqnu.edu.cn

\* Correspondence: sunchengjie@stu.cqut.edu.cn

**Abstract:** The refined segmentation of nuclei and the cytoplasm is the most challenging task in the automation of cervical cell screening. The U-Shape network structure has demonstrated great superiority in the field of biomedical imaging. However, the classical U-Net network cannot effectively utilize mixed domain information and contextual information, and fails to achieve satisfactory results in this task. To address the above problems, a module based on global dependency and local attention (GDLA) for contextual information modeling and features refinement, is proposed in this study. It consists of three components computed in parallel, which are the global dependency module, the spatial attention module, and the channel attention module. The global dependency module models global contextual information to capture a priori knowledge of cervical cells, such as the positional dependence of the nuclei and cytoplasm, and the closure and uniqueness of the nuclei. The spatial attention module combines contextual information to extract cell boundary information and refine target boundaries. The channel and spatial attention modules are used to provide adaption of the input information, and make it easy to identify subtle but dominant differences of similar objects. Comparative and ablation experiments are conducted on the Herlev dataset, and the experimental results demonstrate the effectiveness of the proposed method, which surpasses the most popular existing channel attention, hybrid attention, and context networks in terms of the nuclei and cytoplasm segmentation metrics, achieving better segmentation performance than most previous advanced methods.

**Keywords:** cervical cell segmentation; global dependency; local attention; contextual information; feature refinement



**Citation:** Li, G.; Sun, C.; Xu, C.; Zheng, Y.; Wang, K. Cervical Cell Segmentation Method Based on Global Dependency and Local Attention. *Appl. Sci.* **2022**, *12*, 7742. <https://doi.org/10.3390/app12157742>

Academic Editors: Min Yang, Hao Liu, Shanxiang Chen and Yinong Chen

Received: 24 June 2022

Accepted: 27 July 2022

Published: 1 August 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

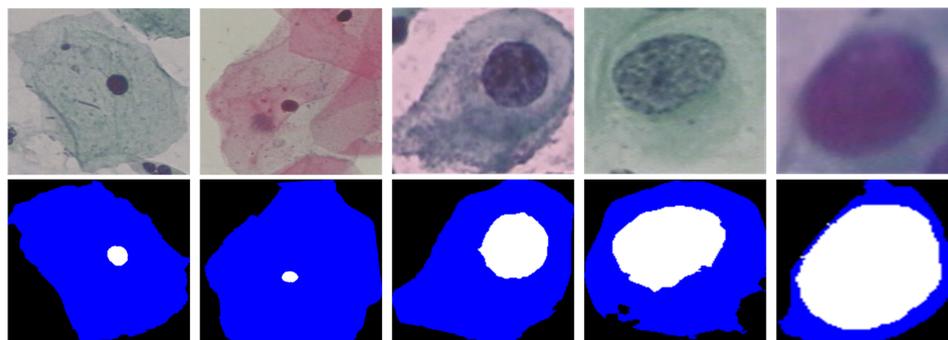
## 1. Introduction

Cervical cancer is one of the most common and deadly cancers among women. Nevertheless, this cancer is completely treatable if detected in the precancerous stage. Cytological examinations, such as the Pap smear or thin-layer cytology, are important routine screening tests for the diagnosis of cervical abnormalities [1]. However, manual analysis is time-consuming, labor-intensive, and error-prone. In addition, due to factors such as the shortage of pathologists and regional economic differences, manual analysis has been unable to meet the urgent needs of an increasing number of patients. In order to automate cervical cell screening, and to improve the accuracy of cervical cytology, computer-aided diagnostic methods based on deep learning have received widespread attention at home and abroad.

The mainstream deep learning-based cervical cell segmentation methods can be broadly classified into two types. The first one addresses the challenge of difficult pixel identification at the boundaries of the cell nuclei and cytoplasm by refining the network segmentation results by using prior knowledge or post-processing [2–6], which guides the model training by prior knowledge and uses the results of model segmentation as input.

The results are optimized using post-processing. Common methods include location-based prior, shape prior, Markov random field, graphical cut [6], edge detection operator, etc. Although this means can optimize the segmentation results, it divides the segmentation task into multiple stages and the detection process is very cumbersome. Another is to improve the segmentation performance by optimizing the network model, and there are relatively few research works of this type. References [7,8] designed complex network models to improve segmentation accuracy, which can achieve end-to-end network training, but is accompanied by heavy computational overhead. In addition, it does not sufficiently take into account a priori knowledge of the location and shape of cervical cells, and the variability of normal and abnormal cells.

Refined nuclear and cytoplasm segmentation is essential for automated cervical cell screening, because when cervical cells are infected, the shape, color, and texture of the nuclei are the first features to appear abnormal, and once the cytoplasmic boundary is located, quantitative evaluation index values such as cell diameter and nucleoplasm ratio can be calculated. As shown in Figure 1, the size, shape, and staining of cervical cells were highly variable. The nuclei of normal cervical cells are tiny, compact, and uniformly colored, while the nuclei of infected cervical cells are swollen and irregular, with blurred borders, low contrast between the target and the background, and significant color differences within the nuclei, making them highly susceptible to false detection. For cervical cells, it is a basic a priori knowledge that the nuclei is only present inside the cell and is closed and unique, regardless of whether the cervical cells are infected or not. The segmentation of cervical cells is a difficult task. To meet these challenges and to exploit the a priori knowledge, a high-performance network for extracting rich features from images is required.



**Figure 1.** The two columns on the left are normal cervical cells and the three columns on the right are infected cervical cells. Normal cervical cells have tiny and compact nuclei. When cervical cells are infected, the nuclei first appear abnormal in shape, color, and texture, among other features.

Recently, U-shaped network structures have been widely used in medical image processing [9–18]. U-Net [18], and its latest extensions, such as Transunet [13], have been the leading medical image segmentation methods in recent years. U-Net, combined with the Transformer [19], has achieved good performance in several medical image tasks; however, in the cervical cell segmentation task, the Transformer is not well applied to this task due to the limitation of data volume [20]. Some works enhance the segmentation performance of classical U-Net networks by designing complex aggregated feature modules between the backbone and segmentation semantic headers, including multi-scale feature integration [21] and multi-level feature integration [10], and moreover, enhanced channel and spatial representations [22,23]. However, both of them do not consider feature differences and information dependency comprehensively, and cannot fully and effectively utilize mixed domain information and contextual information.

Deep learning has been highly successful for computer vision tasks [24], and the ability of network feature extraction is the basis for a variety of downstream tasks. In this study, a network module for contextual information modeling and features refinement based on global dependency and local attention (GDLA) is proposed. For the closed and unique nature of cell nuclei, the global dependency module models contextual information to capture the a priori knowledge of cervical cells. For boundary segmentation challenges, a spatial attention module combined with contextual information is constructed for extracting boundary information to refine target boundaries. The channel and spatial attention modules are used to provide adaption of the input information and make it easy to identify subtle but dominant differences of similar objects. Specifically, the network consists of three main components: the global dependency module, the spatial attention module, and the channel attention module. The global dependency module first uses the weighted average of all location features on the spatial domain of the feature map as the global context features, and then aggregates the global context features to each location feature to enhance the representation of each location feature. The spatial attention module and the channel attention module compress the channel dimension and the spatial dimension, respectively, and use the sigmoid function to activate to get the spatial attention map and the channel attention map. The three modules are computed in parallel and finally applied to the original feature map to complete the adaption of the input information.

The main contributions are summarized as follows:

- A module based on global dependency and local attention for contextual information modeling and feature refinement is proposed, based on a comprehensive consideration of feature differences and information dependency, which can be used as a basic component of the network to effectively enhance feature representation and improve the network's ability to extract and utilize features.
- Considering information dependency, as well as the closure and uniqueness of the nuclei, the global context is modeled using long-range dependency.
- A hybrid attention module is constructed to provide adaptive input information and enhance key information representation to suppress useless information. Among them, the spatial attention module extracts spatial detail information sufficiently and effectively to better refine the target boundaries.
- An improved U-Net network, which implements an end-to-end network training model, is evaluated on the Herlev dataset to show that the segmentation performance can be steadily improved.

## 2. Related Work

### 2.1. Segmentation of Cervical Cells

Accurate segmentation of cervical cancer cells, especially the nuclei, is important for the quantitative analysis of cervical cancer [25]. Traditional image representation-based cervical segmentation methods have been widely used [2–4,26–29]. However, due to the small public dataset of cervical cells and the small amount of data, deep learning has been less studied in this task than in other medical image processing tasks [25,30,31]. For instance, Ref. [4] proposed a learning-based method with a robust shape before segmenting individual cells in a Pap smear image, also incorporating high-level shape information to guide the segmentation. Ref. [29] improved the existing fuzzy c-mean clustering algorithm for nuclear segmentation by finding the optimal number of clusters to segment cell nuclei. Ref. [2] proposed a new superpixel-based Markov random field (MRF) segmentation framework to acquire cell images' nuclei, cytoplasm, and image background. A graph search-based method for improving the segmentation of abnormal cervical nuclei is proposed in [3]. The nuclear shape constraint is embedded in the construction of the segmentation graph. The nuclei of cervical cells were localized and segmented using Mask-RCNN to obtain coarse segmentation results in [32], followed by refining the segmentation using a local fully connected conditional random field. Allehaibi et al. [31] segmented the whole cervical cell image via transfer learning, which was effective but treated the nuclei and cytoplasm

as the same class. A deformable multi-path integration-based model is proposed in [7], which used Dense-Unet as the backbone network and used deformable convolution for nuclei of different shapes and sizes. Ref. [8] proposed a progressive step-by-step training method with small to large image resolutions and shallow to deep networks. A two-stage framework based on Mask-RCNN to automatically segment overlapping cells is proposed in [5]. In the first stage, candidate cytoplasmic bounding boxes are proposed. In the second stage, pixel-to-pixel alignment is used to refine the boundaries. Greenwald et al. [33] generate preliminary predictions for each nucleus and boundary in the image, and then use the predictions of the nucleus and cell boundary as input to the watershed algorithm to create a final instance segmentation mask for each cell and nucleus in the image. In this study, an end-to-end network training model is implemented using a modified U-Net network. It involves capturing global dependency and refining the features to enhance feature representation in the encoder part, and then connecting refined feature maps to the decoder network via skip connection. It can steadily improve network performance and enhance segmentation without additional prior knowledge and post-processing.

### 2.2. Global Dependency

Capturing global dependency, which aims to extract the global understanding of a visual scene, is proven to benefit a wide range of tasks. Recently, global context information is applied to other respective tasks, and this achieved an excellent performance [34–36]. Both convolutional and recurrent operations are constructive blocks that deal with one local neighborhood at a time [37]. To capture long-range dependency in image data, for example, it is a common practice to stack convolutional layers, and as the number of layers deepens and the perceptual field becomes larger, it is possible to incorporate the originally non-adjacent pixel points into an overall consideration, and the breadth of the acquired information distribution becomes higher and higher. The above method of capturing global dependency is not only inefficient, but also makes network optimization difficult. Wang et al. [37] computes the response at a position as a weighted sum of the features at all positions via a non-local operation for capturing global dependency. Cao et al. simplified it by directly modeling the global context as a weighted average of all location features, and then aggregating global context features to the features at each location in [38]. A self-attentive mechanism is introduced in [39] to capture feature dependency in the spatial dimension and channel dimension, respectively. Similar to [37], in order to use the association between any two points of features in space to mutually enhance the representation of the respective features, the association strength matrix between any two points of features needs to be computed, which has the problem of large computational and parametric quantities.

### 2.3. Local Attention

The method of shifting attention to the most important regions of an image and ignoring irrelevant parts is called attention mechanism [40]. Attention mechanisms have been highly successful in both natural language processing and computer vision, and attention-base networks, proposed in [41,42], implicitly and adaptively predict potential key features. An SE (Squeeze and Excitation) block [41] is divided into two parts, a squeeze module and an excitation module. Global spatial information is collected in the squeeze module via global average pooling. The excitation module captures channel-wise relationships and outputs an attention vector by using fully-connected layers and nonlinear layers. A CBAM (convolutional block attention module) [42] sequentially infers attention maps along two separate dimensions, channel and spatial, then the attention maps are multiplied to the input feature map for adaptive feature refinement. Subsequently, much work has been done to improve them, with a more detailed design of the local attention module [43,44]. However, SE and CBAM are still two of the existing widely used attention networks, due to their robustness.

### 3. Proposed Method

The segmentation of cervical cells is a tricky task, due to their large variation in size, shape, and staining, low contrast between target and background, and blurred and irregular boundaries. A high-performance network for extracting rich features from images is required to meet these challenges. In this paper, we propose a module for global information modeling and feature refinement (GDLA), which consists of three main components: a global dependency module, a spatial attention module, and a channel attention module. The global dependency module models the global context information to automatically extract a priori knowledge of the shape and location of cervical cells, such as the closeness and uniqueness of nuclei. The spatial attention module combines contextual information to extract boundary information and refine target boundaries. The channel and spatial attention modules are used to provide adaption of the input information. The three networks are computed in parallel and finally applied to the original feature map to enhance the feature representation. Given an intermediate feature map  $X \in \mathbb{R}^{C \times H \times W}$  as an input, GDLA infers in parallel a global dependency map  $F_G \in \mathbb{R}^{C \times 1 \times 1}$ , a channel attention map  $F_C \in \mathbb{R}^{C \times 1 \times 1}$ , and a spatial attention map  $F_S \in \mathbb{R}^{1 \times H \times W}$ . The overall framework structure of the GDLA Model is shown in Figure 2, which does not change the feature map size and can be easily applied to different layers of the backbone. The overall process can be summarized as:

$$\hat{X} = \{F_G(X) \oplus X\} \oplus \{F_C(X) \otimes X\} \oplus \{F_S(X) \otimes X\} \tag{1}$$

where  $\otimes$  and  $\oplus$  denote element-wise multiplication and element-wise addition. During multiplication or addition, the tensor values are broadcast (copied), accordingly.  $\hat{X}$  is the final refined output. Figure 3 depicts the computation process of each component. The following describes the details of each module.

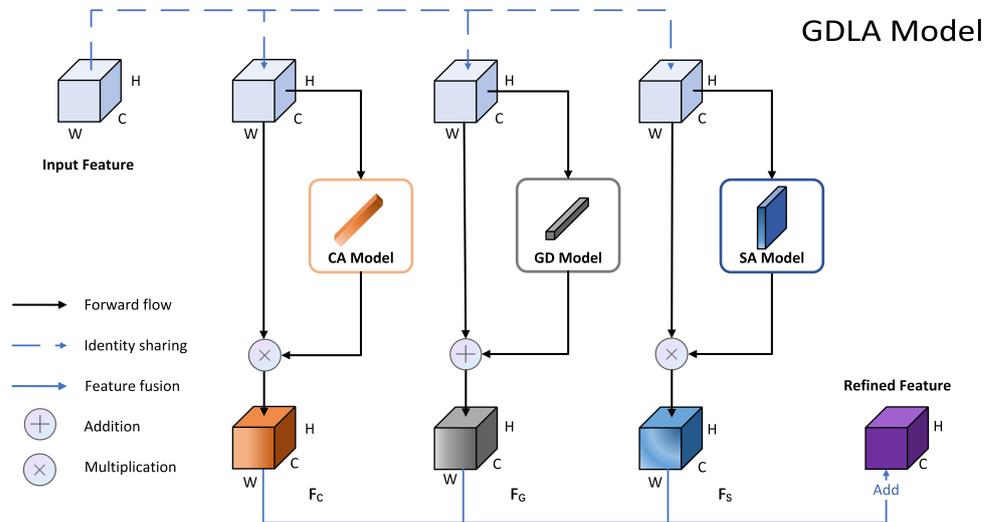
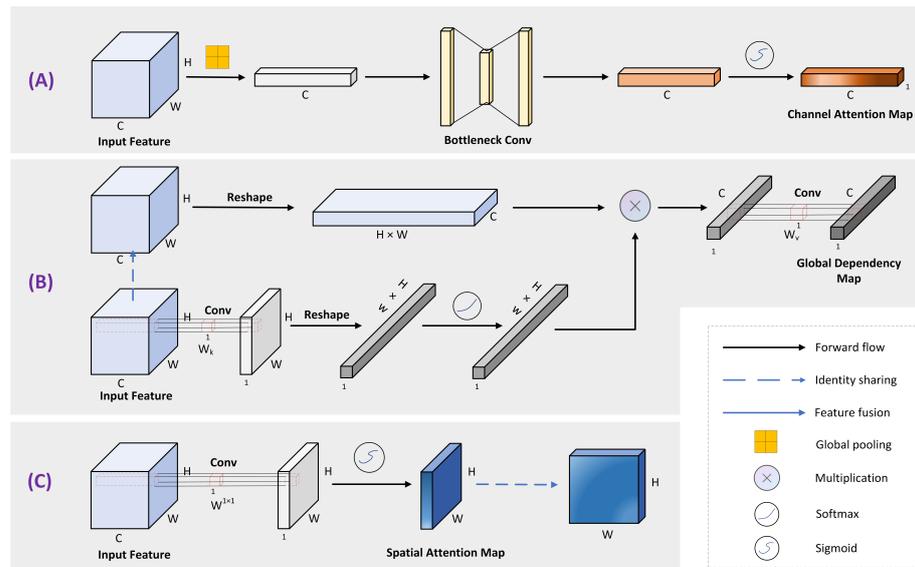


Figure 2. Overview of GDLA model.



**Figure 3.** Diagram of each attention sub-module, where (A–C) denote CA Model, GD Model, and SA Model, respectively.

### 3.1. Global Dependency Module

Both convolutional and recurrent operations are constructive blocks that deal with one local neighborhood at a time. To capture global dependency in image data, for example, the common practice is to stack convolutional layers, and as the layers deepen, the perceptual field becomes larger and larger. It is possible to incorporate the originally non-adjacent pixel points into an overall consideration, and the breadth of the acquired information distribution becomes higher and higher. However, as the network deepens, a series of problems come along with it, such as model degradation, gradient disappearance or explosion, and difficulty in network training. In order to obtain global context information while avoiding these problems, the global dependency module in this study refers to the simplified version of the Non-local Block (NL Block) mentioned in [38]. Denote  $X = \{X_i\}_{i=1}^{N_p}$  as the feature map of one input instance, where  $N_p$  is the number of positions in the feature map.  $N_p = H \times W$ . GD module can be expressed as:

$$F_{G_i}(X_i) = W_v \sum_{j=1}^{N_p} \frac{\exp(W_k X_j)}{\sum_{n=1}^{N_p} \exp(W_k X_n)} X_j \quad (2)$$

where  $i$  is the index of the query location and  $j$  enumerates all possible locations.  $W_v$  and  $W_k$  represent the linear transformation matrices, and  $X$  and  $F_G$  represent the input and output of the GD module, respectively. As shown in Figure 3B, the input feature mapping utilizes the Embedded Gaussian [37] to construct global context information. Specifically, the feature weight map is obtained using  $1 \times 1$  convolution  $W_k$  and the softmax function, and all query locations share an attention map to obtain global context features, which will be  $1/(H \times W)$  of the original computational complexity compared to computing the attention map [40] for each point on the feature map, enabling it to be used as the network base component to be embedded in different network layers. Subsequently, the global context features are feature transformed using  $1 \times 1$  convolution  $W_v$ , and finally the global context features are aggregated to each query position through element-wise addition operation.

### 3.2. Channel Attention Module

A certain intermediate input feature mapping  $X = [x^1, x^2, x^3 \dots, x^c]$  is denoted as a combination of channels  $x^i, x^i \in \mathbb{R}^{H \times W}$ , and channel attention is obtained using a squeeze-excitation operation that transforms the input feature map into  $X_{CA} \in \mathbb{R}^{1 \times H \times W}$  by globally

averaging the squeeze space dimensions as a combination of channel features  $x_{CA}^i \in \mathbb{R}^{1 \times 1}$ . The squeeze operation for each channel can be expressed as:

$$x_{CA}^i(x_i) = 1/(H \times W) \sum_{j=1}^{N_p} x^i(j) \quad (3)$$

After spatial squeezing, the feature vector is reduced by a bottleneck structure to limit the computational complexity of the model. Specifically, a dimensionality reduction operation is first performed on the channel dimension with a rate of  $r$ , plus *ReLU* activation function for nonlinear transformation, and finally expanded to the original dimensionality. Subsequently, the sigmoid function is used to excite the channel so that the dynamic range of the eigenvector  $X_{CA}$  is brought to the interval  $[0 \sim 1]$ . This strategy makes it easy to identify subtle but dominant differences of similar objects. The overall process can be summarized as:

$$F_C(X) = \sigma(W_1(\text{ReLU}(\text{BN}(W_0(X_{CA})))))) \quad (4)$$

where  $\sigma$  denotes the sigmoid function.  $W_0$  and  $W_1$  are two linear transformation matrices,  $W_0 \in \mathbb{R}^{C \times C/r}$  and  $W_1 \in \mathbb{R}^{C/r \times C}$  used to construct the bottleneck, and *BN* is the batch normalization process. The structure diagram of the channel attention module is shown in Figure 3A. The excitation feature vectors represent the importance of the channels used to rescale the original features  $x^i$ , enabling adaptive adjustment of channel weights to emphasize meaningful channels. In this paper, a  $1 \times 1$  convolutional layer is used to construct the bottleneck structure layer, and a *BN* layer is added for regularization after the dimensionality reduction operation using convolution. Since the smaller the channel compression ratio, the more detailed the differentiation of channel importance, the ratio  $r$  is set to 2.

### 3.3. Spatial Attention Module

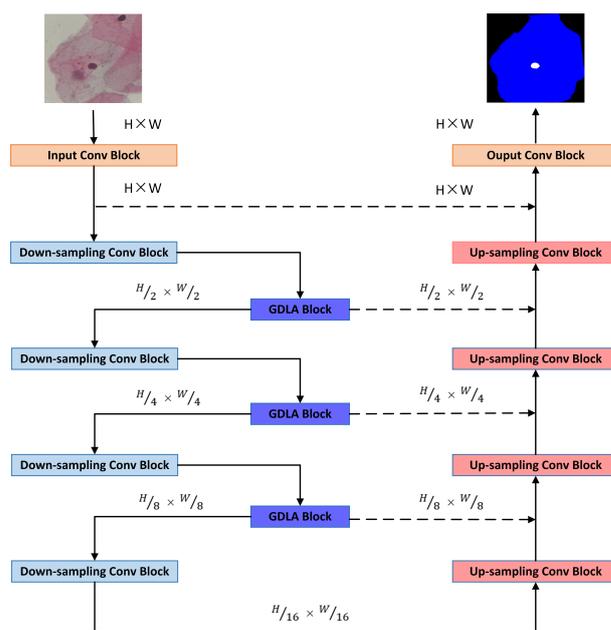
Spatial attention is also obtained through the squeeze-excitation operation. Unlike channel attention, spatial attention compresses the channel dimension of the input feature map  $X$  by using a  $1 \times 1$  convolution layer, and channel squeezing transforms the input feature map into  $X_{SA} \in \mathbb{R}^{1 \times H \times W}$ . The channel-wise information is integrated by this operation, and the features at each spatial location after integration are linear combinatorial representations of the features of all channels corresponding to that point, which are subsequently excited in the spatial dimension. The overall process can be summarized as:

$$F_S(X) = \sigma(W^{1 \times 1}(X)) \quad (5)$$

where  $W^{1 \times 1}$  represents a convolution operation with a filter size of  $1 \times 1$ . The structure of the spatial attention module is illustrated in Figure 3C. The value of each spatial point  $x_{SA}^i$  on the feature map  $F_S$  after excitation is in the interval  $[0 \sim 1]$ , which is used to indicate the importance of spatial information on the feature map, where  $i$  is defined as a certain spatial point on the feature map.

## 4. Experiment

To evaluate the proposed method in this paper, experiments were conducted on the Herlev dataset with the Zijdenbos similarity index (ZSI) as the prime evaluation metric. U-Net was used as a baseline to validate our proposed method in this study. The specific network instance structure is illustrated in Figure 4. The experimental results show that the proposed GDLA module leads to a stable improvement in network performance. It surpasses the most popular existing channel attention networks, hybrid attention networks, and contextual networks in terms of the nuclei and cytoplasm segmentation metrics; achieves better segmentation performance than most previous advanced methods; and demonstrates its ability as a fundamental component of the network that can stably improve the network's ability to extract and utilize feature information.



**Figure 4.** Illustration of the network structure used with the GDLA module embedded in U-Net.

#### 4.1. Dataset and Implementation Details

In this study, a series of experiments are implemented on the Herlev dataset. This dataset was published by Herlev University Hospital. They collected a total of 917 color images of cervical cells, including two main categories: normal cervical cells and abnormal cervical cells. The abnormal cervical cells were divided into four subcategories: carcinoma in situ cells (150 cell images), moderate dysplastic cells (146 cell images), severe dysplastic cells (197 cell images), and light dysplastic cells (182 cell images), for a total of 675 images of abnormal cervical cells. Normal cervical cells were divided into three subclasses, corresponding to normal columnar cells (98 cell images), normal intermediate cells (70 cell images), and normal superficial cells (74 cell images), for a total of 242 diseased cervical cells. These images were collected with different equipment and staining techniques and manually labeled by a professional pathologist. The manual labeling was divided into four categories, namely, background, nucleus, cytoplasm, and unknown regions; unknown regions were considered as background in this study.

The Herlev dataset is divided into 8:2, which means that 734 images are utilized for training, and 183 images are set for testing. Notably, the data are divided separately according to cervical cell types. For example, the carcinoma in situ cells contain a total of 150 images, of which 120 images are randomly selected as the training set, and the others are used as the test set. All images are scaled to a size of  $256 \times 256$  pixels. In addition, to enhance the model generalization, some general online data augmentations are utilized in the training phase, including flip, rotation, random affine transform, random brightness contrast, and image blur.

The deep learning framework is PyTorch 1.7, with the experimental code in Python 3.8. Additionally, the experiment is implemented on a PC (Windows 10 operating system, Intel(R) Xeon(R) Gold 6226R CPU @ 2.90 GHz, 16 Gbyte RAM, and NVIDIA Quadro RTX 5000 GPU). The model optimizer uses SGD with a batch size of 16 and an initial learning rate of  $1 \times 10^{-4}$ .

The model is trained for a total of 200 epochs, with the “poly” learning rate strategy, in which the initial rate is multiplied by  $(1 - \text{iter}/\text{max\_iter})^{\text{power}}$  for each iteration, with power 0.9. The traditional cross-entropy loss function is used and the parameters are kept consistent for all experiments. Unless otherwise noted, the experiments in this work will be conducted with the channel compression ratio set to 2. To ensure the reliability of the results, five trials are performed for each experiment, and then the mean  $\mu$  and standard deviation  $\sigma$  of the results are taken. Moreover, all experiments employ the same random seed.

#### 4.2. Evaluation Metrics

To evaluate the segmentation performance of the model, three evaluation metrics are used in this paper to assess the performance of the model. These three evaluation metrics are widely used in the cervical cell segmentation task, which are pixel-based *Precision*, *Recall*, and the Zijdenbos similarity index (*ZSI*), respectively. The evaluation metrics are calculated as follows:

$$\text{Precision} = TP / (TP + FP) \quad (6)$$

$$\text{Recall} = TP / (TP + FN) \quad (7)$$

$$\text{ZSI} = 2 \cdot TP / (2 \cdot TP + FP + FN) \quad (8)$$

where *TP* indicates the number of positive pixel samples (i.e., foreground) that are correctly classified, *TN* indicates the number of negative pixel samples (i.e., background) that are correctly classified, *FP* means the number of positive pixel samples that are misclassified, and *FN* is the number of negative pixel samples that are misclassified.

#### 4.3. Comparison Experiments

The performance differences between the proposed method and the mainstream channel attention network [41], hybrid attention network [42], and contextual network [38] for cervical cell nuclei and cytoplasm segmentation were compared in the first place.

The performance of the modules was comparatively compared by embedding GC (Global Context) Block [38], SE Block [41], CBAM Block [42], and GDLA Block in U-Net, where GC, SE, and CBAM have been described in Section 2. Table 1 shows that the proposed method outperforms the other three networks, both for nuclei segmentation and cytoplasm segmentation. This is due to the fact that our proposed method integrates local feature differences and global information dependency, and effectively captures the closure and uniqueness of nuclei, refines cell boundaries, and provides adaptive input information by modeling contextual information and refining features. Compared with the baseline, GDLA has a higher performance (approximately 3.8% ↑ on ZSI of nuclei and 3.2% ↑ on ZSI of cytoplasm) with a slight increase in #params. Figure 5 shows the visual results of the effect of different modules. In Figure 5, the bar graph indicates the results of evaluation metrics for nucleus segmentation, and the line graph indicates the results of evaluation metrics for cytoplasm segmentation.

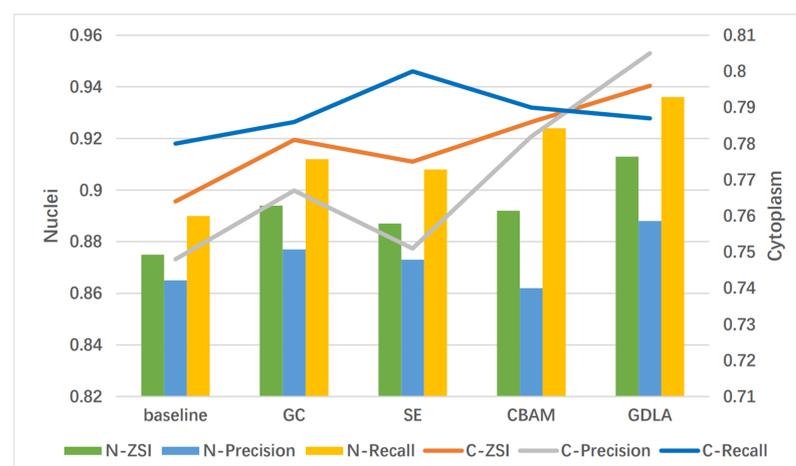


Figure 5. Visual results of the effect of different modules.

**Table 1.** The effect of different modules, including GC unit, SE unit, CBAM module, and GDLA module, on network performance. The values are in the format of  $\mu \pm \sigma$ .

Methods	Nuclei			Cytoplasm			#Params
	ZSI	Precision	Recall	ZSI	Precision	Recall	
UNet (baseline)	0.875 $\pm$ 0.12	0.865 $\pm$ 0.13	0.890 $\pm$ 0.13	0.764 $\pm$ 0.14	0.748 $\pm$ 0.12	0.780 $\pm$ 0.15	13.40 M
UNet + GC [38]	0.894 $\pm$ 0.11	0.877 $\pm$ 0.13	0.912 $\pm$ 0.12	0.781 $\pm$ 0.13	0.767 $\pm$ 0.13	0.786 $\pm$ 0.12	13.74 M
UNet + SE [41]	0.887 $\pm$ 0.10	0.873 $\pm$ 0.10	0.908 $\pm$ 0.12	0.775 $\pm$ 0.12	0.751 $\pm$ 0.12	0.800 $\pm$ 0.13	13.44 M
UNet + CBAM [42]	0.892 $\pm$ 0.10	0.862 $\pm$ 0.09	0.924 $\pm$ 0.12	0.786 $\pm$ 0.13	0.782 $\pm$ 0.15	0.790 $\pm$ 0.15	13.48 M
UNet + GDLA (ours)	0.913 $\pm$ 0.09	0.888 $\pm$ 0.11	0.936 $\pm$ 0.13	0.796 $\pm$ 0.11	0.805 $\pm$ 0.08	0.787 $\pm$ 0.11	14.09 M

Moreover, the proposed GDLA is compared with the dedicated networks on nucleus segmentation metrics, since most of the dedicated networks target only the segmentation of cervical nuclei. Table 2 shows the experimental results of several methods that have been specially designed for extracting the features of cervical cells on the Herlev dataset. The results are all from the original papers [2,7,8,26,27,29,30,32,45]. As can be seen, the proposed method achieves the competitive ZSI score on cervical cell nuclei segmentation and demonstrates an outstanding segmentation performance on the Herlev dataset. Among them, although Gap-search MRF and PUG-Net have achieved comparable results to the proposed method, the former uses the traditional method of Markov random field for the pixel-level classification of cervical cell images, which divides the cervical cell segmentation task into multiple steps and makes the process tedious and complicated. The method uses a non-local mean filter to attenuate the “noise”, generates superpixels, and manually extracts a 13-dimensional feature vector for each superpixel. Finally, a solving algorithm is required. There is instability in long assembly lines and complex processes. Any segmentation errors in the intermediate steps can lead to the failure of the entire segmentation process. The latter, both reference [7] and reference [8], on the other hand, although using a deep learning approach, utilizes the idea of the model ensemble to incrementally train the neural network. Furthermore, although reference [32] achieves the highest ZSI score on cervical cell nuclei segmentation, its method only focuses on singlet nuclei segmentation, meaning that Mask-RCNN is utilized to locate the nuclei before segmentation, which to some extent excludes background interference while ignoring the important qualitative and quantitative indicators of cervical cell characteristics such as the nucleoplasmic ratio. Compared to these methods, the proposed approach is more simple, efficient, and elegant.

**Table 2.** Comparison of our proposed method with state-of-the-art methods specially designed for cervical cell nuclei segmentation. The values are in the format of  $\mu \pm \sigma$ .

Methods	ZSI	Precision	Recall
RGVF algorithm [26]	0.87 $\pm$ 0.19	0.83 $\pm$ 0.20	0.96 $\pm$ 0.13
Unsupervised [27]	0.89 $\pm$ 0.15	0.88 $\pm$ 0.15	0.93 $\pm$ 0.15
FCM [29]	0.80 $\pm$ 0.24	0.85 $\pm$ 0.21	0.83 $\pm$ 0.25
SP-CNN [30]	0.90 $\pm$ N/A	0.89 $\pm$ N/A	0.91 $\pm$ N/A
Gap-search MRF [2]	0.91 $\pm$ 0.07	N/A	N/A
Mask-RCNN [32]	0.95 $\pm$ 0.10	0.96 $\pm$ 0.05	0.96 $\pm$ 0.11
D-MEM [7]	0.933 $\pm$ 0.14	0.946 $\pm$ 0.06	0.984 $\pm$ 0.00
PUG-Net [8]	0.911 $\pm$ 0.10	0.890 $\pm$ 0.12	0.950 $\pm$ 0.11
HMLS [45]	0.90 $\pm$ 0.12	0.91 $\pm$ 0.13	0.93 $\pm$ 0.13
Proposed	0.913 $\pm$ 0.09	0.888 $\pm$ 0.11	0.936 $\pm$ 0.13

#### 4.4. Ablation Experiments

The ablation experiments are divided into two parts, which are to verify the effectiveness of GDLA embedding and to module parallel computing.

The GDLA Block is gradually added to U-Net to verify its effectiveness in enhancing the network’s ability to extract and utilize feature information. The GDLA is proposed to be embedded in the U-Net encoder section after each stage, and then the refined feature maps are connected to the decoder network by skip connection. Table 3 shows the results

of integrating the GDLA block at different stages. Furthermore, the results are visualized in Figure 6 for easy observation. The results show that the network performance is steadily improved when the GDLA block is added gradually. When embedding the GDLA module only in s1, the model is comparable to the #params and performance of embed GC block in Table 1, but with both fewer blocks embedded and #params (approximately 0.3 million ↓).

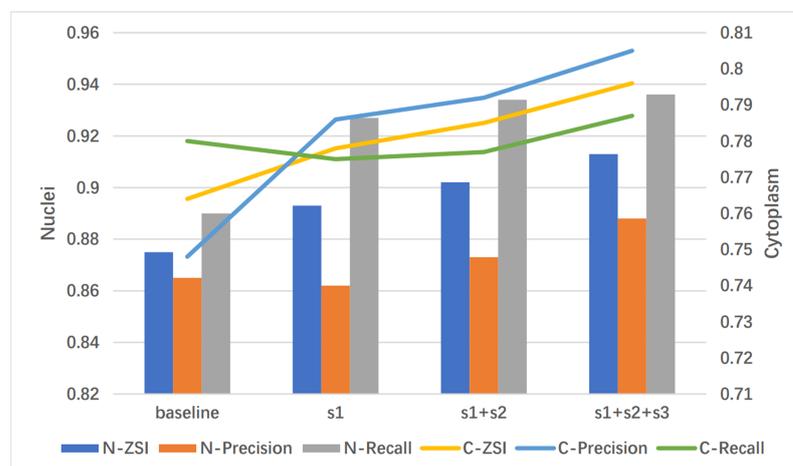
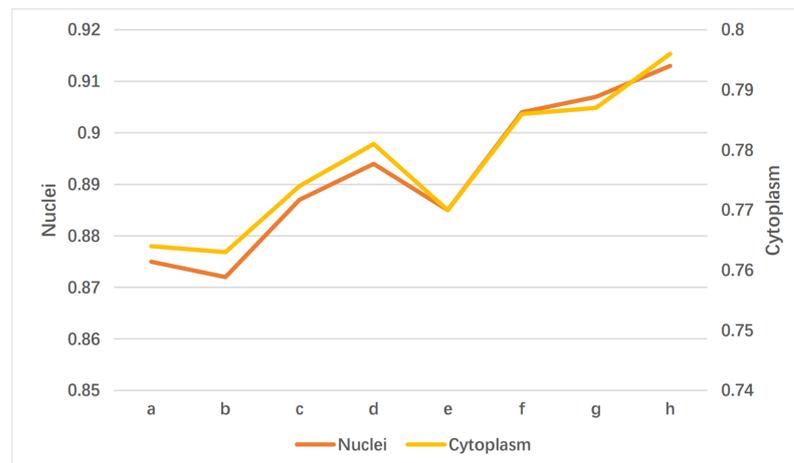


Figure 6. Visual results of embedding the GDLA module at different stages of U-Net.

Table 3. The results of embedding the GDLA module at different stages of U-Net. The values are in the format of  $\mu \pm \sigma$ .

Methods	Nuclei			Cytoplasm			#Params
	ZSI	Precision	Recall	ZSI	Precision	Recall	
baseline	0.875 ± 0.12	0.865 ± 0.13	0.890 ± 0.13	0.764 ± 0.14	0.748 ± 0.12	0.780 ± 0.15	13.40 M
+ s1	0.893 ± 0.11	0.862 ± 0.12	0.927 ± 0.14	0.778 ± 0.11	0.786 ± 0.12	0.775 ± 0.12	13.43 M
+ s1 + s2	0.902 ± 0.11	0.873 ± 0.11	0.934 ± 0.12	0.785 ± 0.10	0.792 ± 0.12	0.777 ± 0.12	13.56 M
+ s1 + s2 + s3	0.913 ± 0.09	0.888 ± 0.11	0.936 ± 0.13	0.796 ± 0.11	0.805 ± 0.08	0.787 ± 0.11	14.09 M

Table 4 shows the effects of the three components of the GDLA network on the network performance, and the model feasibility of their combined computation. Nuclei and cytoplasm segmentation were evaluated using several metrics, with ZSI as the main evaluation metric. Here, only the changes in the ZSI evaluation indexes are shown in Table 4, to facilitate the observation of the comparison result. From Table 4, it can be seen that the GD module alone has a greater improvement on the network than the other two modules (approximately 1.9% ↑ for nuclei and 1.7% ↑ for cytoplasm). The ZSI increases by 1.2% for nuclei and 1.0% for cytoplasm, compared to the baseline when using the CA module alone. Interestingly, the segmentation performance of the network is slightly reduced when using SA alone (approximately 0.3% ↓ on nuclei). However, when SA and GD are considered as a combination, the performance of the network is both improved, with a slightly better performance when using the combination of GD and SA than when using the combination of GD and CA (approximately 0.3% ↑ on nuclei). The best network performance is achieved when all three modules are applied simultaneously. Figure 7 shows the visual results of disassembling the GDLA module using the evaluation metrics ZSI. The categories ‘a–h’ at the bottom of Figure 7 correspond to the eight combination cases in Table 4. For instance, ‘a’ represents the original baseline without any embedded modules, and ‘h’ represents all three modules being applied.



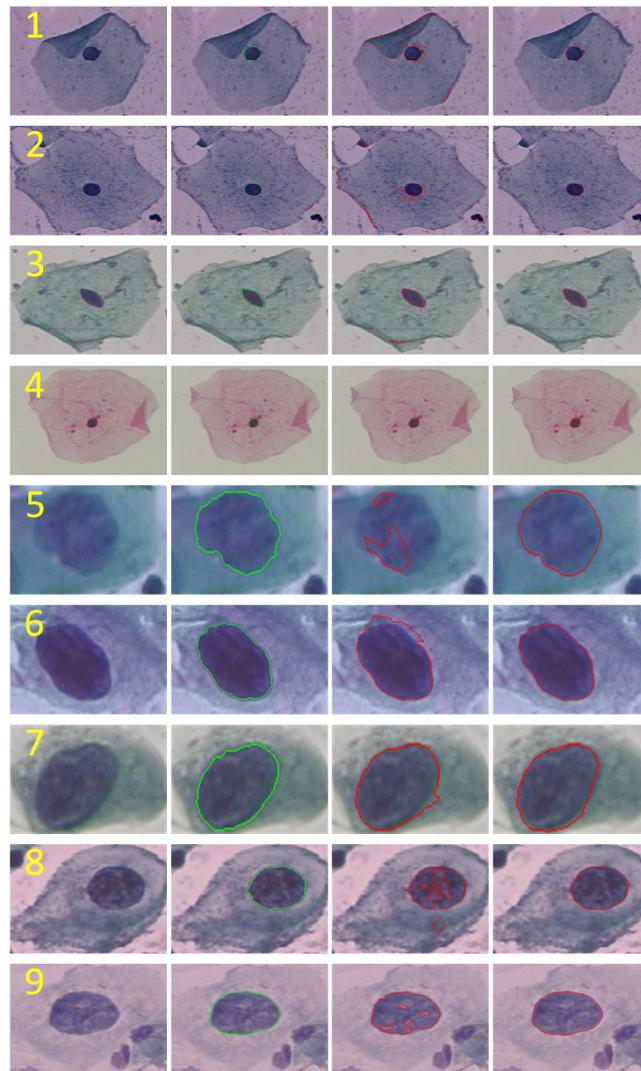
**Figure 7.** Visual results of disassembling the GDLA module.

**Table 4.** Disassembling GDLA module. The effects of the three components of the GDLA network on the network performance. The values are in the format of  $\mu \pm \sigma$ .

SA	CA	GD	Nuclei	Cytoplasm
✓			0.875 ± 0.12	0.764 ± 0.14
	✓		0.872 ± 0.08	0.763 ± 0.10
		✓	0.887 ± 0.10	0.774 ± 0.10
✓	✓		0.894 ± 0.07	0.781 ± 0.08
	✓	✓	0.885 ± 0.12	0.770 ± 0.11
✓	✓	✓	0.904 ± 0.10	0.786 ± 0.15
✓		✓	0.907 ± 0.12	0.787 ± 0.13
✓	✓	✓	0.913 ± 0.09	0.796 ± 0.11

#### 4.5. Subjective Effect Analysis

In addition to the objective metrics, this paper also shows the subjective effect of segmenting some images in the Herlev dataset. As shown in Figure 8, the results were obtained from segmenting some images in the Herlev dataset using the U-Net and GDLAU-Net networks, where GDLAU-Net represents the network based on the U-Net framework after embedding the GDLA Block. The first four rows are normal cervical cells, and the last five rows are abnormal cervical cells. As shown in Figure 8, the GDLAU-Net network has better segmentation performance than U-Net for both normal and abnormal cells. From 1, 2, 3, and 8 in Figure 8, it can be seen that U-Net is susceptible to noise and background color, leading to multiple and wrong detections due to the lack of global contextual semantic information. The GDLAU-Net has excellent robustness for this. From 4 in Figure 8, it can be seen that GDLAU-Net is better for the segmentation of small targets, compared to U-Net. From 5 in Figure 8, it can be seen that GDLAU-Net is better than U-Net for the segmentation of cells with blurred boundary contours after expansion. From 6 and 7 in Figure 8, it can be seen that GDLAU-Net is finer for cell nucleus boundary segmentation compared with U-Net, due to the full extraction and utilization of spatial information. From 8 and 9 in Figure 8, it can be seen that GDLAU-Net can better capture the global information dependence and achieve the closure and uniqueness of cell nuclei segmentation.



**Figure 8.** Visual segmentation results. From left to right: pap smear images, manual annotations, U-Net segmentation effect, and GDLAU-Net segmentation effect.

## 5. Conclusions

In this study, we proposed a module based on global dependency and local attention, for contextual information modeling and feature refinement (GDLA). It captures global information dependency and takes full advantage of the closed and complete, as well as unique properties of cervical cells. Moreover, combining local feature attention can effectively enhance feature representation and elegantly improve the network's ability to extract and utilize features. An end-to-end network for the cervical cell nucleus and cytoplasm segmentation was implemented by embedding it into multiple layers of the U-Net network. The GDLA module has been shown to steadily improve the network performance, surpassing the most popular channel attention, hybrid attention, and contextual networks, and achieving a better segmentation performance than most previous advanced methods.

In future work, we will not only compute in parallel, but we will explore the influence of the arrangement compute form of modules in GDLA on cervical cell segmentation, and extend GDLA to other deep learning algorithms and other medical imaging tasks.

**Author Contributions:** Conceptualization, C.S. and G.L.; methodology, C.S.; validation, C.S., C.X., K.W. and Y.Z.; formal analysis, C.S.; investigation, C.S.; resources, C.S.; data curation, C.S.; writing—original draft preparation, C.S.; writing—review and editing, G.L., C.S., C.X. and Y.Z.; visualization,

C.S. and K.W.; supervision, G.L. and C.X.; project administration, G.L. and C.S.; funding acquisition, G.L. and C.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by the China Chongqing Science and Technology Commission under Grant cstc2020jscx-msxmX0086, cstc2019jscx-zdztzx0043, and cstc2019jcyj-msxmX0442; the China Chongqing Banan District Science and Technology Commission project under Grant 2020QC413; and China Chongqing Municipal Education Commission under Grant KJQN202001137. Moreover, this work is from the Chongqing University of Technology Graduate Education Quality Development Action Plan Funding Results (Project number: gzlcx20223456).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data are contained within the article and anyone can be use these data.

**Conflicts of Interest:** The authors declare no conflicts of interest.

### Abbreviations

The following abbreviations are used in this manuscript:

GDLA	Global Dependency and Local Attention
MRF	Markov random field
ZSI	Zijdenbos similarity index
SA	Spatial Attention
CA	Channel Attention
GD	Global Dependency

### References

- Cohen, P.A.; Jhingran, A.; Oaknin, A.; Denny, L. Cervical cancer. *Lancet* **2019**, *393*, 169–182. [[CrossRef](#)]
- Zhao, L.; Li, K.; Wang, M.; Yin, J.; Zhu, E.; Wu, C.; Wang, S.; Zhu, C. Automatic cytoplasm and nuclei segmentation for color cervical smear image using an efficient gap-search MRF. *Comput. Biol. Med.* **2016**, *71*, 46–56. (In English) [[CrossRef](#)] [[PubMed](#)]
- Zhang, L.; Kong, H.; Liu, S.; Wang, T.; Chen, S.; Sonka, M. Graph-based segmentation of abnormal nuclei in cervical cytology. *Comput. Med. Imaging Graph.* **2017**, *56*, 38–48. (In English) [[CrossRef](#)] [[PubMed](#)]
- Song, Y.; Tan, E.L.; Jiang, X.; Cheng, J.-Z.; Ni, D.; Chen, S.; Lei, B.; Wang, T. Accurate Cervical Cell Segmentation from Overlapping Clumps in Pap Smear Images. *IEEE Trans. Med. Imaging* **2017**, *36*, 288–300. [[CrossRef](#)] [[PubMed](#)]
- Chen, J.; Zhang, B. Segmentation of Overlapping Cervical Cells with Mask Region Convolutional Neural Network. *Comput. Math. Methods Med.* **2021**, *2021*, 3890988. (In English) [[CrossRef](#)] [[PubMed](#)]
- Song, Y.; Zhang, L.; Chen, S.; Ni, D.; Lei, B.; Wang, T. Accurate Segmentation of Cervical Cytoplasm and Nuclei Based on Multiscale Convolutional Network and Graph Partitioning. *IEEE Trans. Biomed. Eng.* **2015**, *62*, 2421–2433. [[CrossRef](#)]
- Zhao, J.; Li, Q.; Li, X.; Li, H.; Zhang, L. Automated Segmentation of Cervical Nuclei in Pap Smear Images Using Deformable Multi-Path Ensemble Model. In Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 8–11 April 2019; pp. 1514–1518.
- Zhao, J.; Dai, L.; Zhang, M.; Yu, F.; Li, M.; Li, H.; Wang, W.; Zhang, L. PGU-net+: Progressive Growing of U-net+ for Automated Cervical Nuclei Segmentation. In *Multiscale Multimodal Medical Imaging*; Springer: Cham, Switzerland, 2020; pp. 51–58.
- Jha, D.; Smedsrud, P.H.; Johansen, D.; de Lange, T.; Johansen, H.D.; Halvorsen, P.; Riegler, M.A. A Comprehensive Study on Colorectal Polyp Segmentation With ResUNet++, Conditional Random Field and Test-Time Augmentation. *IEEE J. Biomed. Health* **2021**, *25*, 2029–2040. [[CrossRef](#)]
- Stoyanov, D.; Taylor, Z.; Carneiro, G.; Syeda-Mahmood, T.; Martel, A.; Maier-Hein, L.; Tavares, J.M.R.; Bradley, A.; Papa, J.P.; Belagiannis, V.; et al. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. In Proceedings of the 4th International Workshop (DLMIA 2018) and 8th International Workshop (ML-CDS 2018), Granada, Spain, 18–20 September 2018; pp. 3–11.
- Kim, T.; Lee, H.; Kim, D. UACANet: Uncertainty Augmented Context Attention for Polyp Segmentation. In Proceedings of the 29th ACM International Conference on Multimedia, Chengdu, China, 20–24 October 2021; pp. 2167–2175.
- Srivastava, A.; Jha, D.; Chanda, S.; Pal, U.; Johansen, H.D.; Johansen, D.; Riegler, M.A.; Ali, S.; Halvorsen, P. Msrf-net: A multi-scale residual fusion network for biomedical image segmentation. *arXiv* **2021**, arXiv:2105.07451.
- Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv* **2021**, arXiv:2102.04306.
- Valanarasu, J.M.J.; Sindagi, V.A.; Hacihaliloglu, I.; Patel, V.M. KiU-Net: Overcomplete Convolutional Architectures for Biomedical Image and Volumetric Segmentation. *IEEE Trans. Med. Imaging* **2022**, *41*, 965–976. [[CrossRef](#)]

15. Jha, D.; Riegler, M.A.; Johansen, D.; Halvorsen, P.; Johansen, H.D. DoubleU-Net: A Deep Convolutional Neural Network for Medical Image Segmentation. In Proceedings of the 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS), Rochester, MN, USA, 28–30 July 2020; pp. 558–564.
16. Zhou, H.-Y.; Guo, J.; Zhang, Y.; Yu, L.; Wang, L.; Yu, Y. nnFormer: Interleaved Transformer for Volumetric Segmentation. *arXiv* **2021**, arXiv:2109.03201.
17. Valanarasu, J.M.J.; Patel, V.M. UNExT: MLP-based Rapid Medical Image Segmentation Network. *arXiv* **2022**, arXiv:2203.04967.
18. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Munich, Germany, 5–9 October 2015; pp. 234–241.
19. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16 × 16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
20. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 9992–10002.
21. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6230–6239.
22. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention u-net: Learning where to look for the pancreas. *arXiv* **2018**, arXiv:1804.03999.
23. Roy, A.G.; Navab, N.; Wachinger, C. Concurrent Spatial and Channel ‘Squeeze & Excitation’ in Fully Convolutional Networks. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2018, Granada, Spain, 16–20 September 2018; pp. 421–429.
24. Liu, Z.; Mao, H.; Wu, C.-Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A ConvNet for the 2020s. *arXiv* **2022**, arXiv:2201.03545.
25. Rahaman, M.M.; Li, C.; Wu, X.; Yao, Y.; Hu, Z.; Jiang, T.; Li, X.; Qi, S. A Survey for Cervical Cytopathology Image Analysis Using Deep Learning. *IEEE Access* **2020**, *8*, 61687–61710. [[CrossRef](#)]
26. Li, K.; Lu, Z.; Liu, W.; Yin, J. Cytoplasm and nuclei segmentation in cervical smear images using Radiating GVF Snake. *Pattern Recogn.* **2012**, *5*, 1255–1264. [[CrossRef](#)]
27. Gençtav, A.; Aksoy, S.; Önder, S. Unsupervised segmentation and classification of cervical cell images. *Pattern Recogn.* **2012**, *45*, 4151–4168. [[CrossRef](#)]
28. Chankong, T.; Theera-Umpon, N.; Auephanwiriyakul, S. Automatic cervical cell segmentation and classification in Pap smears. *Comput. Methods Programs Biomed.* **2014**, *113*, 539–560. (In English) [[CrossRef](#)]
29. Sharma, B.; Mangat, K.K. An improved nucleus segmentation for cervical cell images using FCM clustering and BPNN. In Proceedings of the 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, India, 21–24 September 2016; pp. 1924–1929.
30. Gautam, S.; Bhavsar, A.; Sao, A.K.; Harinarayan, K.K. CNN based segmentation of nuclei in PAP-smear images with selective pre-processing. In *Medical Imaging 2018: Digital Pathology*; International Society for Optics and Photonics: Houston, TX, USA, 2018; p. 105810X.
31. Allehaibi, K.H.S.; Nugroho, L.E.; Lazuardi, L.; Prabuwo, A.S.; Mantoro, T. Segmentation and classification of cervical cells using deep learning. *IEEE Access* **2019**, *7*, 16925–116941.
32. Liu, Y.; Zhang, P.; Song, Q.; Li, A.; Zhang, P.; Gui, Z. Automatic segmentation of cervical nuclei based on deep learning and a conditional random field. *IEEE Access* **2018**, *6*, 53709–53721. [[CrossRef](#)]
33. Greenwald, N.F.; Miller, G.; Moen, E.; Kong, A.; Kagel, A.; Dougherty, T.; Fullaway, C.C.; McIntosh, B.J.; Leow, K.X.; Schwartz, M.S.; et al. Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nat. Biotechnol.* **2022**, *40*, 555–565. (In English) [[CrossRef](#)]
34. Cao, Z.; Huang, Z.; Pan, L.; Zhang, S.; Liu, Z.; Fu, C. TCTrack: Temporal Contexts for Aerial Tracking. *arXiv* **2022**, arXiv:2203.01885.
35. Konwer, A.; Xu, X.; Bae, J.; Chen, C.; Prasanna, P. Temporal Context Matters: Enhancing Single Image Prediction with Disease Progression Representations. *arXiv* **2022**, arXiv:2203.01933.
36. Rao, Y.; Zhao, W.; Chen, G.; Tang, Y.; Zhu, Z.; Huang, G.; Zhou, J.; Lu, J. DenseCLIP: Language-Guided Dense Prediction with Context-Aware Prompting. *arXiv* **2021**, arXiv:2112.01518.
37. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-local neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 7794–7803.
38. Cao, Y.; Xu, J.; Lin, S.; Wei, F.; Hu, H. GCNet: Non-Local Networks Meet Squeeze-Excitation Networks and Beyond. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Korea, 27–28 October 2019; pp. 1971–1980.
39. Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; Lu, H. Dual attention network for scene segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 3146–3154.
40. Guo, M.-H.; Xu, T.-X.; Liu, J.-J.; Liu, Z.-N.; Jiang, P.-T.; Mu, T.-J.; Zhang, S.-H.; Martin, R.R.; Cheng, M.-M.; Hu, S.-M. Attention Mechanisms in Computer Vision: A Survey. *arXiv* **2021**, arXiv:2111.07624.

41. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
42. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
43. Wang, Q.L.; Wu, B.G.; Zhu, P.F.; Li, P.; Zuo, W.; Hu, Q. ECA-net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 11531–11539.
44. Misra, D.; Nalamada, T.; Arasanipalai, A.U.; Hou, Q.B. Rotate to attend: Convolutional triplet attention module. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 3–8 January 2021; pp. 3138–3147.
45. Braga, A.M.; Marques, R.C.; Medeiros, F.N.; Neto, J.F.R.; Bianchi, A.G.; Carneiro, C.M.; Ushizima, D.M. Hierarchical median narrow band for level set segmentation of cervical cell nuclei. *Measurement* **2021**, *176*, 109232. [[CrossRef](#)]