



OPEN ACCESS

EDITED BY

Guichao Li,
Fudan University, China

REVIEWED BY

Habib Izadkhah,
University of Tabriz, Iran
Mohamed Elbadawy,
Benha University, Egypt

*CORRESPONDENCE

Ruirong Tan

✉ ruirongtanlab@163.com

Pengfei Yu

✉ pfyu@ynu.edu.cn

†These authors share first authorship

SPECIALTY SECTION

This article was submitted to
Genitourinary Oncology,
a section of the journal
Frontiers in Oncology

RECEIVED 08 October 2022

ACCEPTED 06 February 2023

PUBLISHED 24 April 2023

CITATION

Zhang S, Li L, Yu P, Wu C, Wang X, Liu M,
Deng S, Guo C and Tan R (2023) A deep
learning model for drug screening and
evaluation in bladder cancer organoids.
Front. Oncol. 13:1064548.
doi: 10.3389/fonc.2023.1064548

COPYRIGHT

© 2023 Zhang, Li, Yu, Wu, Wang, Liu, Deng,
Guo and Tan. This is an open-access article
distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

A deep learning model for drug screening and evaluation in bladder cancer organoids

Shudi Zhang^{1†}, Lu Li^{2†}, Pengfei Yu^{1*}, Chunyue Wu²,
Xiaowen Wang¹, Meng Liu², Shuangsheng Deng²,
Chunming Guo² and Ruirong Tan^{3*}

¹School of Information Science and Engineering, Yunnan University, Kunming, China, ²College of Life Sciences, Yunnan University, Kunming, China, ³Center for Organoids and Translational Pharmacology, Translational Chinese Medicine Key Laboratory of Sichuan Province, Sichuan Institute for Translational Chinese Medicine, Sichuan Academy of Chinese Medicine Sciences, Chengdu, China

Three-dimensional cell tissue culture, which produces biological structures termed organoids, has rapidly promoted the progress of biological research, including basic research, drug discovery, and regenerative medicine. However, due to the lack of algorithms and software, analysis of organoid growth is labor intensive and time-consuming. Currently it requires individual measurements using software such as ImageJ, leading to low screening efficiency when used for a high throughput screen. To solve this problem, we developed a bladder cancer organoid culture system, generated microscopic images, and developed a novel automatic image segmentation model, AU2Net (Attention and Cross U2Net). Using a dataset of two hundred images from growing organoids (day1 to day 7) and organoids with or without drug treatment, our model applies deep learning technology for image segmentation. To further improve the accuracy of model prediction, a variety of methods are integrated to improve the model's specificity, including adding Grouping Cross Merge (GCM) modules at the model's jump joints to strengthen the model's feature information. After feature information acquisition, a residual attentional gate (RAG) is added to suppress unnecessary feature propagation and improve the precision of organoids segmentation by establishing rich context-dependent models for local features. Experimental results show that each optimization scheme can significantly improve model performance. The sensitivity, specificity, and F1-Score of the ACU2Net model reached 94.81%, 88.50%, and 91.54% respectively, which exceed those of U-Net, Attention U-Net, and other available network models. Together, this novel ACU2Net model can provide more accurate segmentation results from organoid images and can improve the efficiency of drug screening evaluation using organoids.

KEYWORDS

deep learning, bladder cancer organoids, image segmentation, drug screening, U2Net model

1 Introduction

Bladder cancer is a malignant tumour that occurs on the bladder mucosa and is a common malignancy of the urinary system. In recent years, the incidence of bladder cancer has risen rapidly, especially in men and the elderly population. According to the latest data from GLOBOCAN, bladder cancer ranks 13th in the ranking of malignant tumors, and its incidence rate ranks among the top 10 (1). Most patients with early bladder cancer are unaware of their condition and have no apparent symptoms, resulting in a low early diagnosis rate (2). Therefore, early detection and diagnosis are crucial for bladder cancer management.

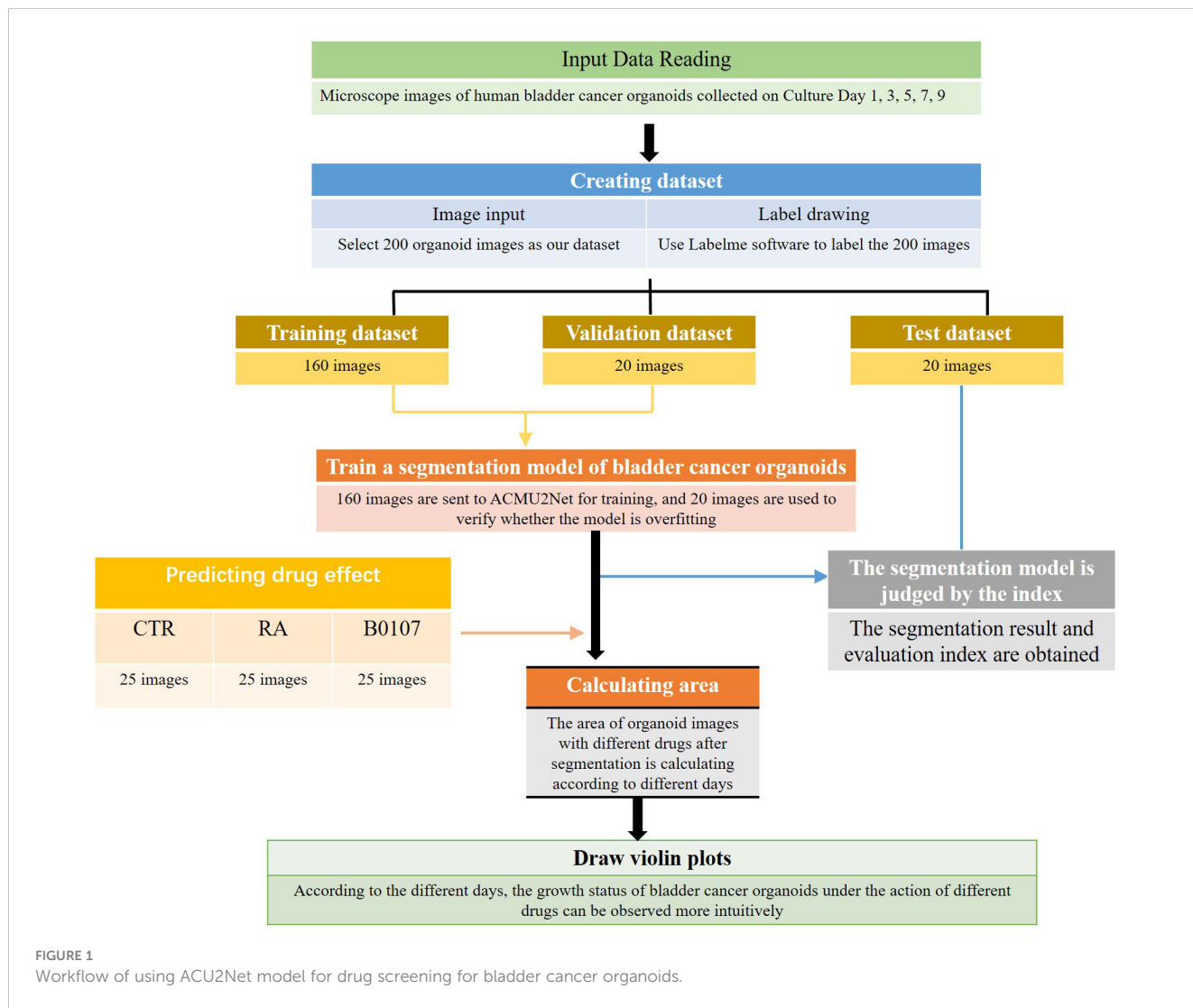
An organoid (3–5) is a multicellular *in vitro* tissue cultured in three dimensions (3D). It can mimic the original tissue environment and is similar to the primary organ in terms of physical and chemical properties. Compared with cell lines and animal experiments currently and commonly conducted, organoid experiments have a broader application prospect in various preclinical and clinical studies. Human bladder cancer patients' organoids had been established and exome sequencing confirmed that the spectrum of genetic mutation from cultured organoids is heterogeneous and may represent bladder cancer tumour evolution *in vivo* (6), in addition, patient derived organoids recapitulate the histopathological diversity of human bladder cancer. Therefore, bladder organoids could be used for drug screening and could serve as a preclinical platform for precision medicine. Indeed, the drug response using patient derived tumor organoids show partial correlations with mutational profiles and changes associated with drug resistance, especially the response can be validated in xenografts model (7), furthermore, the drug screen could be easily performed *in vitro* rapidly to match cancer patient's unique mutation. Beside bladder cancer, 3D cell culture and organoid approaches are being increasingly used for other cancer types and even broader scenario in basic research and drug discovery in a variety of diseases (8). Recent advances in these technologies, which enable the *in vitro* study of these tissue-like structures, are very important for assessment of heterogenous diseases, such as colorectal cancer (9). As well, the use of 3D culture will not only help advance basic research but also help reduce the use of animals in biomedical science (10). 3D organoid methods are also very important for the study of diseases in conditions when it is difficult to obtain patient tissues. Therefore, developing methods to rapidly and accurately process organoid image data become a top priority. Accelerating the analysis of organoid image data is also extremely important for the development of biological experiments. A bladder cancer organoid derived from a patient is an *in vitro* model of their bladder tumour, and it can be used as a model to rapidly select for the most appropriate therapeutic strategies for each individual patient (6, 11, 12). The great advantage of the bladder cancer organoid is that this model is an accurate mimic of the bladder cancer patient's tumour (7). At present, the role of the drug sensitivity prediction model in the precision treatment of bladder cancer has gradually become prominent. This model is able to determine the most sensitive drugs for patients *in vitro* and evaluate

their therapeutic effects. The introduction of organoid models into the diagnosis and treatment of bladder cancer not only predicts resistance to anti-cancer drugs but also identifies effective cancer therapy for individual patients (13, 14).

Deep learning is a machine learning method based on artificial neural networks with representation learning. By using rapid and continuous automatic analysis of pathological images, the deep learning method could dramatically improve the accuracy rate and efficiency of drug screening and evaluation. Despite the rapid development of deep learning applications in pathological images, such as retinal segmentation (15), and pulmonary nodule segmentation (16), development of a deep learning segmentation method suitable for cancer organoids, especially for evaluation of bladder cancer organoids and drug screening, requires more efforts and is less studied.

This paper proposes a new Attention and Cross U2Net (ACU2Net) model for bladder cancer organoid segmentation in images. We used an image segmentation neural network in deep learning for bladder cancer organoid segmentation. First, 200 images SW780 and T24 were selected as our data set. Since image segmentation is mainly done by supervised training, our data set has no labels. Therefore, we used Labelme to label the obtained images with masks and outline each organoid. Then, combined with the data characteristics that our organoid background region is more than the target region, we optimized the original U2Net network model and selected a better loss function to improve the segmentation accuracy of the model. After segmentation, the trained model was further used to quickly calculate the area of bladder cancer organoids treated with or without different drugs on different culture days. To quantify the growth of organoids, violin plots were generated permitting the easy visualization of the effect of drug treatment on different days. The workflow for this process is shown in Figure 1.

Our innovations are that this is the first time the image segmentation network has been applied in the application of bladder cancer organoid segmentation. Our model also added a grouping and merging module to the original U2Net (17) jump connection. This combination enriched the characteristic information of bladder cancer organoids, especially the semantic and spatial characteristics. An improved attention mechanism (18) was introduced to make the feature extraction network pay more attention to key features in the feature layer and spatial region of bladder cancer organoids while ignoring unnecessary features in redundant regions. After segmentation, the computer is used to automatically determine the growth of organoids treated with different drugs or different drugs at different days, and then generate violin diagrams for faster and easier visualization for drug screening. The ACU2Net model was developed and tested on 3D bladder cancer organoids for drug screening and evaluation. Our results show that the ACU2Net model is faster, as traditional manual bladder organoid drug screening takes 2-3 days, while the deep learning algorithm only takes about 1 hour and is more accurate. In summary, our method can provide more accurate segmentation results for organoid images and greatly improve the efficiency of drug screening evaluation.



2 Related work

2.1 Clinical trial simulation

Currently, drug screening strategies have been developed based both on experimental and virtual methods. Experimental screening refers to screening in the laboratory. Clinical trial simulation, a virtual screening method, has recently emerged as an interdisciplinary subject and attracted widespread attention in the pharmaceutical industry (19–21). Statistics show that during the stages of new drug development, 60% of drugs fail due to poor drug metabolism or high toxicity. The emergence of clinical trial simulation can guide experimental design in drug research, by conducting computer simulations on key hypotheses before bench experiments, and can potentially obtain drug effect information. Therefore, clinical trial simulation can reduce research investment in clinical trials and improve research efficiency and success rates (22).

Since the discovery of new drug targets and action mechanisms has become more complex, drug screening nowadays requires a huge investment of time and money. To solve this problem, many pharmaceutical companies and drug research institutions have

applied artificial intelligence, such as deep learning methods (23), to improve screening efficiency and to mine for new findings from existing data. For example, DOCK Blaster (24), was developed by the Shoichet Laboratory for molecular docking. It screens in a specific database with a given receptor structure to search for potential active small molecules. However, the parameter settings for this method are troublesome and often influence the accuracy and reliability of the results. Iscreen (25), developed by the Taiwan YC Laboratory, can conduct virtual screening of traditional Chinese medicines online, but only on a small scale, with low diversity and a long screening time. DeepScreen (26) uses convolutional neural networks to obtain single-cell images based on flow-cytometry cells. Compared with standard experimental methods, DeepScreen can significantly reduce the detecting time, from a few days to 2–6 hours, effectively improving the detecting efficiency.

2.2 Image segmentation network

The concept of deep learning originated from the research of artificial neural networks. The structure of deep learning is a multi-

layer perceptron structure with multiple hidden layers. It combines low-level features and then forms more abstract high-level features to represent attribute categories or features. Deep learning theory includes many different deep neural network models (27), such as classic deep neural network (DNN), convolutional neural network (CNN), deep Boltzmann machine (DBM), and recurrent neural network (RNN). Networks with different structures are suitable for processing different data types. For example, CNN is suitable for image processing, and RNN is suitable for speech recognition. At the same time, some different variants of these network models will be produced by combining them with different algorithms.

Deep learning has significantly progressed in image classification (28), image segmentation (29), and object detection (30). Deep learning has also been gradually applied in the field of biomedical image segmentation in natural image segmentation. For example, one group (31) used a fully convolutional network to solve the task of cell image segmentation. The network can classify each pixel and achieve end-to-end training, but the pooling operation causes a slight loss of adequate information, which is less effective in fine-grained segmentation. Another group (32) proposed U-Net, to compensate for information loss through deconvolution layers and feature stitching. It has a simple structure, few parameters, and strong plasticity. As a result, it is one of the most basic and effective models currently applied to cell image segmentation. However, the network is not practical for edge contour segmentation, because there are many cell breakage problems and the depth of the network that can be constructed is limited due to vanishing gradients problems. To improve the U-Net model, some deep learning networks have been proposed. For instance, UNet++ (33) connects encoders and decoders using dense skip connections between different layers. In addition, FU-Net (34) uses a dynamically weighted cross-entropy loss to improve U-Net. Quan et al. (35) combined the features extracted by the segmentation network to build a more profound network architecture to achieve more accurate cell segmentation. Moreover, attention mechanisms have recently been used in image segmentation to improve the segmentation effect. For example, Fu et al. (36) proposed a dual attention network attention mechanism that captures contextual information dependencies through a self-attention mechanism and adaptively integrates local features and global information. However, efficiency was only mildly improved through this model, as well, in this model there are too many parameters and it takes too much run time.

3 Methods

Given the characteristics of bladder cancer organoids in microscope images, this paper introduces Grouping Cross Merge (GCM) in the skip connection of the U2Net network on the basis of U2Net and Attention U-Net (37) and combines GCM with the encoder and decoder of the U2Net network. The shallow and deep features are cross-merged by the GCM module. In addition, to further enhance the correlation between features, a feature transfer channel is added between each GCM module, so that the features of the entire U2Net network are serially passed to the last GCM

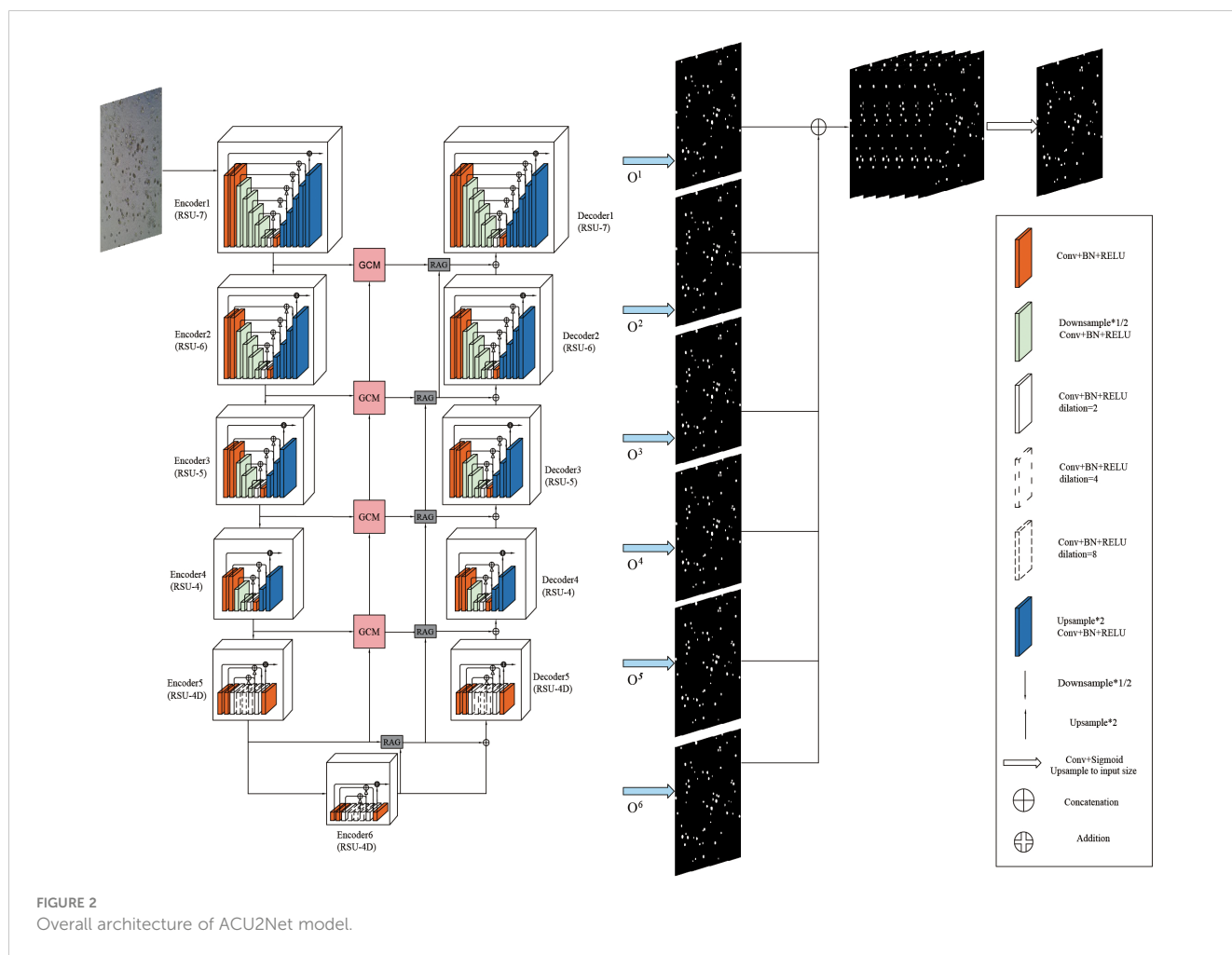
module and output. Next, it is passed through the Residual Attention Gate (RAG), which enhances the model's ability to extract the features of the region of interest. Finally, bladder cancer organoids and backgrounds are classified through a 1×1 convolutional layer. The proposed model is shown in Figure 2.

3.1 Residual U module

The residual U-module consists of three parts: an input layer that collects local features and transforms channels, a U-shaped structure that extracts and encodes multi-scale context information, and an output layer that fuses the input and intermediate layers. First, the convolutional layer, which is a standard for local feature extraction, converts the input feature $x(H \times W \times C_{in})$ map into an intermediate $F_1(x)$ map with C_{out} channels. Next, a U-shaped symmetric encoder-decoder structure with a height of L, which takes the intermediate feature map $U(F_1(x))$ as input, learns to extract and encode multi-scale context information, and finally fuses the residual connections of local features and multi-scale features by summing. The U-shaped symmetric codec structure representation is shown in Figure 3, where L is the number of layers in the encoder, C_{in} and C_{out} represents the number of input and output channels, and M represents the number of channels in the RSU middle layer. The more significant L is, we can get a deeper residual U-shaped block (RSU), more pooling operations, more diversity of the receptive field, and richer local and global features. Therefore, we can configure the level parameter L to extract multi-scale features from input feature maps with arbitrary spatial resolution. In this paper we selected $L=7$. When using larger L, the sampling of the feature graph will lead to the loss of useful context, resulting in higher computing and memory costs. Using a smaller L less local semantic information can be captured. In contrast, $L=7$ can capture global semantic information and local semantic information and consume fewer computing resources and computing costs.

3.2 The overall structure of the U2Net module

The overall structure of the U2Net network includes 6 encoders and 5 decoders, as well as a saliency map fusion module connected to the decoder. The configured RSU-L fills each stage, and the level parameter L is configured according to the spatial resolution of the input feature map. As can be seen from Figure 2, the left part of the network is the down-sampling process. The first 4 stages fill Encoder1, Encoder2, Encoder3, and Encoder4 with RSU, whose level parameters L are 7, 6, 5, and 4, respectively. For feature maps with larger resolutions, we use a more significant L to obtain information at a larger scale, and the size of the feature maps in each stage is halved layer by layer and restored to the original size. The feature maps in Encoder5 and Encoder6 are of relatively low resolution, and further down-sampling of these feature maps will result in the loss of adequate contextual feature information.



Therefore, RSU-4D is used in both stages, where “D” indicates that RSU is a dilated version in which we replace the pooling and up-sampling operations with dilated convolutions, which means that all intermediate feature maps of RSU-4D have the same resolution as its input feature maps. The last two stages are filled by RSU configured with dilated convolution, and the size of the feature graph within the stage remains unchanged. In the encoding stage, the RSUs are connected by a 2x2 max pooling, and the feature map size becomes 1/32 of the original. The right side of the network is a decoding module composed of 5 decoders, which is an up-sampling process. The RSU configuration in each stage is the same as the left symmetrical position, so the decoder has a similar structure to its symmetric encoder. In Decoder5, we also use an extended version of the residual U-block RSU-4D, similar to that used in encoder-level Encoder5 and Encoder6. The input for each decoder level is a decoder output from its previous level and a symmetric encoder output. In addition, the decoding part changes the size of the reduced feature map to the original size.

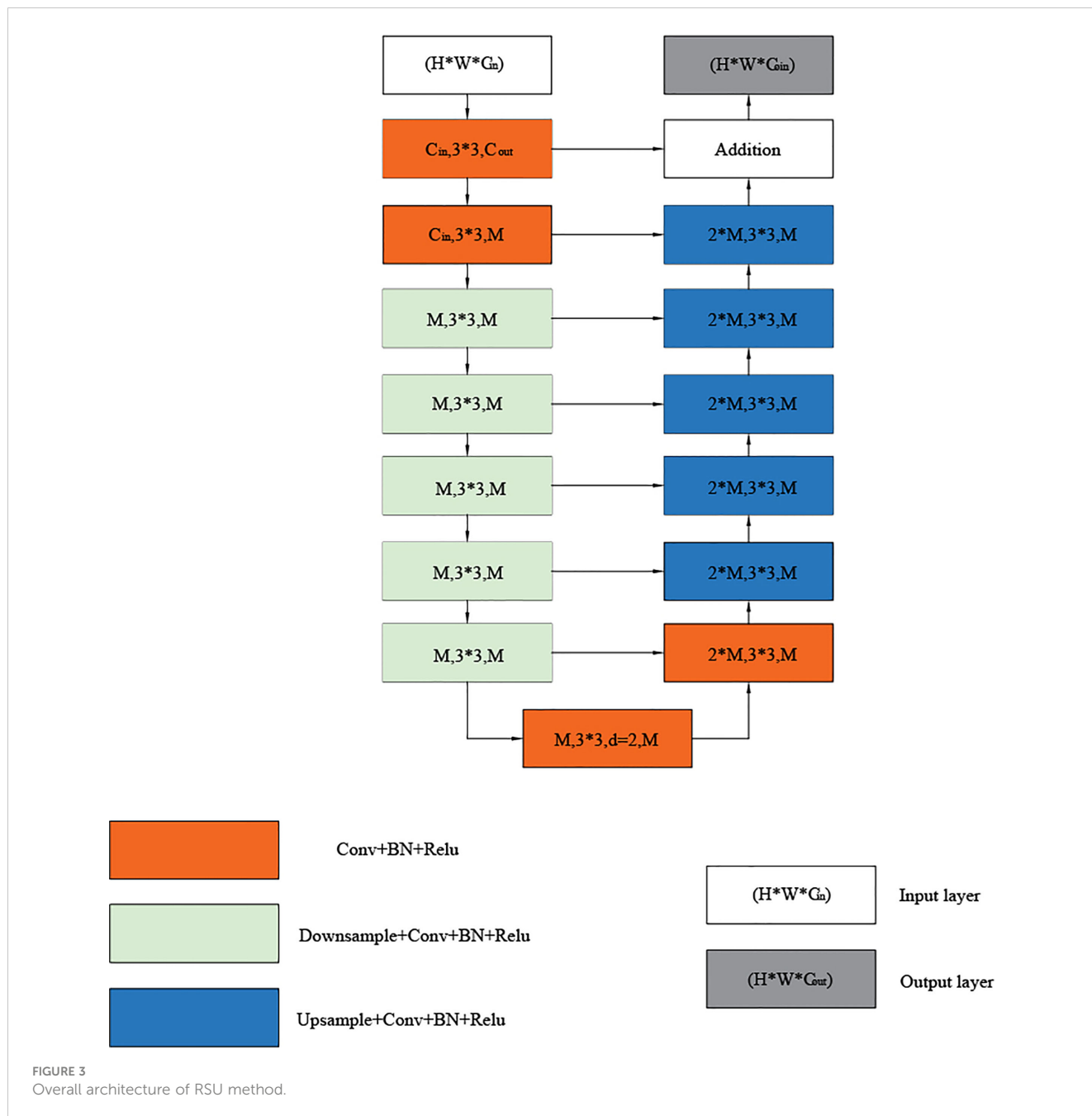
The last part is the feature fusion module, which generates segmentation probability maps. U2Net first generates six output feature maps O^6 , O^5 , O^4 , O^3 , O^2 , and O^1 from Encoder6, Decoder5, Decoder4, Decoder3, Decoder2, and Decoder1 through a 3x3 convolutional layer and a sigmoid function. Then, it up-samples the side output feature map to the input image size, and the six output

feature maps are spliced to generate the final segmentation map. The network structure parameters of U2Net are shown in Table 1.

3.3 Grouping cross fusion module

In this design, to further improve the correlation of feature information, the Grouping Cross Merge (GCM) is applied in the U2Net network, which fuses low-level morphological features and high-level semantic features to enrich the feature information. The network structure of the packet cross-fusion module designed in this paper is shown in Figure 4.

There are two input feature maps for the GCM module: the “mirror” feature maps for the shallow network and the deep network feature maps for the U2Net network. Each feature map is divided into n groups. Because the channel number of the shallow network is relatively small, and the channel number of the deep network is relatively large, we set the values of n as 4, 6, and 8 according to the different layers of the network, shallow layer, middle layer, and deep layer, respectively. Based on the granularity that needs to be subdivided, the two groups of features are merged using the CONCAT layer. After merging, the feature maps are integrated using a 1 x 1 convolutional layer, and the number of channels of the feature map is halved. At this stage the feature cross-

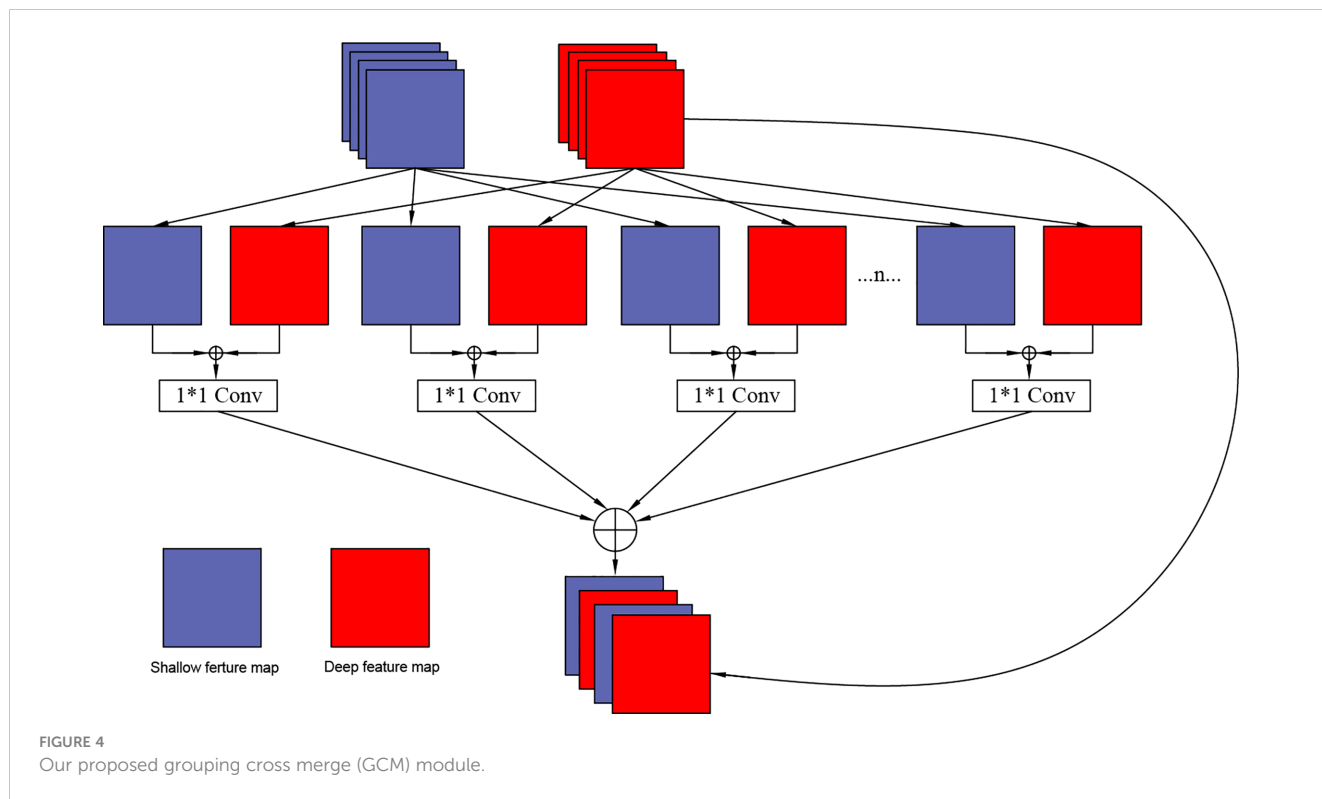


processing of each group is completed. Next, the CONCAT layer is used to combine again to complete the cross- of the two feature maps. Since many convolution calculations are involved in the entire module, to prevent problems such as vanishing gradient, the input feature maps in the deep network are mapped to the cross-merge modules to construct the residual channels of the overall module.

3.4 Residual attention gate

In the Residual Attention Gate (RAG), the feature map of the same layer in the contraction path and the feature map of the

previous layer in the expansion path achieve multi-scale feature fusion through convolution, addition, and activation operations. Then, 1×1 convolution filtering is performed on the fused features, and global average pooling (GAP) is used to extract global features to reduce dimensionality (38). Next, two fully connected layers (FC) are combined with the sigmoid function to generate an organoid-oriented weight matrix, which is then multiplied by the feature fusion graph to adjust the feature weights (39). Since the weight matrix continuously assigns weights to the target area, the value of the background area in the output feature map will be smaller, and the value of the target area will be more significant. Finally, the feature fusion map is residually connected with the adjusted weight feature map to



obtain the final output features. The overall structure of the RAG is shown in Figure 5.

First, the feature maps $x \in \mathbb{R}^{C_s \times W \times H}$ and $g \in \mathbb{R}^{C_g \times W/2 \times H/2}$ are converted from different scale layers into a unified size and mapped

to the feature spaces of a, b , respectively. Then, a 1×1 convolution is performed, and a bit-by-bit summation is carried out. Finally, the fused feature map is obtained by activating the Relu function. The feature map can be defined as ϕ :

TABLE 1 Network parameters of U2Net.

	Block	input size (C × H × W)	output size (C × H × W)	Note
Decoder and Encoder Part	Encoder1	3×256×256	3×256×256	Max Pooling
	Encoder2	64×128×128	128×128×128	Max Pooling
	Encoder3	128×64×64	256×64×64	Max Pooling
	Encoder4	256×32×32	512×32×32	Max Pooling
	Encoder5	512×16×16	512×16×16	Max Pooling
	Encoder6	512×8×8	512×8×8	Dilated Conv
	Decoder5	512×16×16	512×16×16	Dilated Conv
	Decoder4	512×32×32	256×32×32	Up-sampling
	Decoder3	256×64×64	128×64×64	Up-sampling
	Decoder2	128×128×128	64×128×128	Up-sampling
	Decoder1	64×256×256	32×256×256	Up-sampling
Fusion Part of Feature Figure	O^6	512×8×8	1×8×8	Conv+Sigmod
	O^5	512×16×16	1×16×16	Conv+Sigmod
	O^4	256×32×32	1×32×32	Conv+Sigmod
	O^3	128×64×64	1×64×64	Conv+Sigmod
	O^2	64×128×128	1×128×128	Conv+Sigmod
	O^1	32×256×256	1×256×256	Conv+Sigmod

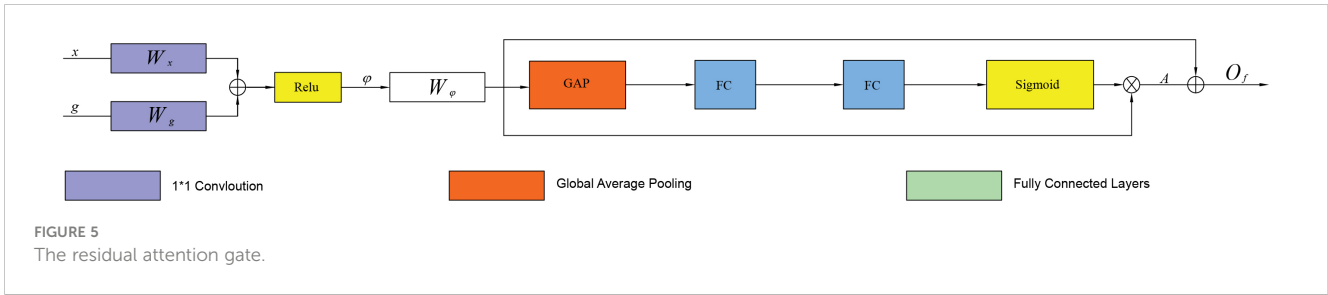


FIGURE 5
The residual attention gate.

$$\varphi = \text{Relu}(a(x) + b(\text{upsampling}(g))) \in \mathbb{R}^{C \times H \times W} \quad (1)$$

In the formula, $a(x) = W_x x$, $b(g) = W_g g$, then the feature φ is sent to the convolutional layer to filter the multi-scale features to generate another feature space d , where $d(\varphi) = W_\varphi \varphi$. W_x , W_g and W_φ represent the 1×1 convolution operation. Next, the GAP layer is applied for dimensionality reduction and global feature extraction to convert it into a one-dimensional vector $a(d(\varphi)) \in \mathbb{R}^{1 \times C}$. Afterwards, two fully connected layers, $f_{c1 \times c/r}$ and $f_{c1 \times c}$ are used to model the correlation between channels (the first layer has c/r channels, the second layer has c channels, and r is the reduction ratio), and outputs the same number of weights as the input features. This operation has more nonlinearity, so that it can better fit the complex correlation between channels, and significantly reduce the number of parameters and computation. Then a Sigmoid function is applied to obtain weights normalized between 0 and 1 to generate a channel attention map A :

$$A = \text{sigmoid}(f_{c1 \times c}(\text{Relu}(f_{c1 \times c/r}(b(d(\varphi)))))) \quad (2)$$

For each channel A_k , $k \in \{1, \dots, c\}$ in the attention map $A \in \mathbb{R}^{1 \times C}$, the weight of the feature map $d(\varphi) \in \mathbb{R}^{C \times H \times W}$ is adjusted. The above calculation retains the useful semantic and detail features of different layers, and the complex background is suppressed as much as possible. Finally, we perform the residual operation, adding pixels with the fusion feature graph to obtain the final output feature o_f :

$$o_f = \sum_{k=1}^c A_k d(\varphi_k) + d(\varphi) \quad (3)$$

3.5 Loss function

The segmentation of bladder cancer organoids is a pixel-level binary classification problem, and we adopt a binary cross-entropy function (40) as part of the loss function. Its mathematical expression is:

$$L_{BCE} = -\frac{1}{n} \sum (y_n \times \ln x_n + (1 - y_n) \times \ln(1 - x_n)) \quad (4)$$

Among them, $y_n = \{0, 1\}$ represents the true value of the dataset, and $x_n = \{0, 1\}$ represents the prediction result.

In medical image analysis, the Dice coefficient is often used as a criterion for judging segmentation results. The Dice loss function proposed based on the Dice coefficient can show whether the

prediction results and the data manually labelled by experts show a consistent distribution in the overall performance (41). At the same time, the Dice loss function focuses on mining the foreground area's information during the network training process. We take the Dice loss function as another part of the loss function, whose mathematical expression is as follows:

$$L_{Dice} = 1 - \frac{2 \times |Pred \cap Label|}{|Pred \cup Label|} \quad (5)$$

Where $Pred$ denotes the predicted image of the network and $Label$ denotes the standard ground truth image.

The overall loss function of the ACU2Net network combines the binary cross-entropy loss function and the Dice loss function with the following mathematical expressions.

$$L_{loss} = L_{BCE} + L_{Dice} \quad (6)$$

4 Experimental results

4.1 Dataset and experimental setup

The dataset used in this study is the image of bladder cancer organoids provided by the Life Science Center of Yunnan University. The data set included 200 images from bladder cancer organoids. The pixel size is 2592×1944 . SW780 and T24 bladder cancer cell lines were used to establish the organoid model (Figure 6). SW780 cells came from low malignancies female patient representing bladder cancer clinical stage I (T1), while T24 cells came from high malignancies female patient representing bladder cancer clinical stage 3 or above. In brief bladder cancer cells were digested by 0.125% or 0.25% trypsin EDTA (Viva Cell, C3530, China) for 3-5 minutes, after 1200 rpm centrifuge, bladder cancer cells were counted and diluted into 250 cells/ μ l. The Cultrex Reduced Growth Factor Basement Membrane Extract, Type 2, PathClear (BME, R&D, 3533-010-02, USA) was prepared on ice, and then added 4 μ l resuspended bladder cancer cells into 40uL BME on ice and mixed gently, finally loaded into a well of 24 wells plate in 37C for 10 minutes. Once BME solidified, 600uL complete DMEM cell culture medium was added into 24 wells plate, and medium were replaced in every 3 days. Organoid images were collected by Microscope DM2000 (Leica, German) at day 1, 3, 5 and 7. The complete culture medium for SW780 bladder cancer organoids was composed of RPMI Medium 1640 basic (Gibco, C118755008T, USA), 10% Fetal Bovine Serum (FBS,

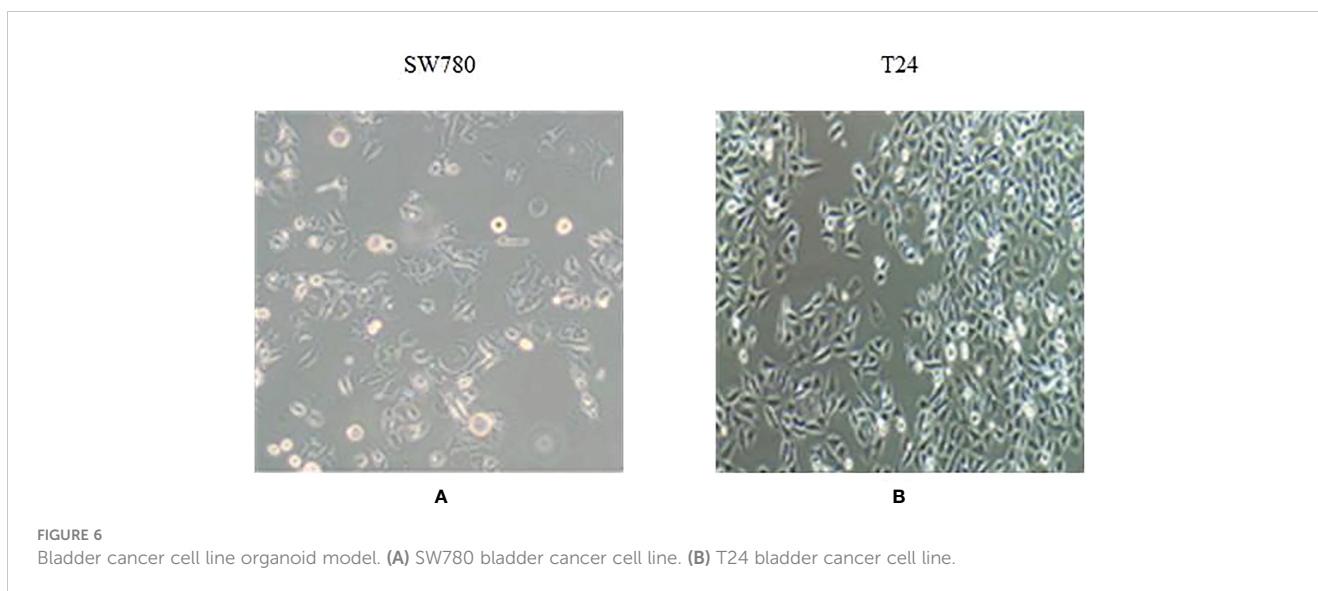


FIGURE 6

Bladder cancer cell line organoid model. (A) SW780 bladder cancer cell line. (B) T24 bladder cancer cell line.

Gibco, 10099141C, USA), 1% Penicillin-Streptomycin (PS, Millipore, TMS-AB2-C, USA), 1% HEPES (Viva Cell, C3544-0100, China), while the medium for T24 cell organoids contained McCoy's 5A Medium (Viva cell, c3020-0500, China), 10% Fetal Bovine Serum (FBS, Gibco, 10099141C, USA), 1% Penicillin-Streptomycin (PS, Millipore, TMS-AB2-C, USA). Since the drug was diluted by DMSO (sigma, D2650, China), the medium of CTR group contained the same proportion of DMSO as the control, while the medium of RA (MedChemExpress, HY-14649, China) and B0107 (MedChemExpress, HY-B0107, China) groups contained RA and B0107. Because the images were collected at different times and from different environments, there exists uneven staining and illumination in the images, therefore it is necessary to denoise them. For this purpose, we chose the non-local average denoising algorithm (42) for processing. Each image is a ground truth image marked by experts at the pixel level. An example of the images in the dataset and its pre-processing are shown in Figure 7.

In order to make full use of the limited dataset, the denoised dataset was expanded by random flipping, cropping, and scaling. Therefore, the 200 images of bladder cancer organoid were expanded to 6400 bladder cancer organoid images with a size of 256×256 pixels. Moreover, the dataset was randomly divided into a training set, a validation set, and a test set in a ratio of 8:1:1, respectively. The experimental environment configuration required for experiments is shown in Table 2.

A total of 1000 epochs were trained, and the gradient descent was optimized using the SGD (Stochastic Gradient Descent) algorithm (43). In addition, loss functions use Dice Loss and BCE Loss, and the neuron inactivation probability was set to 0.5 (44), which is used to further reduce the overfitting phenomenon.

4.2 The evaluation index

The purpose of bladder cancer organoid segmentation is to divide the pixels in the organoid image into organoid pixels and background pixels. There are four possible segmentation results:

True Positive (TP), which indicates that the pixels in the organoid images marked by experts are classified correctly as organoids; False Negative (FN), indicating that the organoid pixels in the images marked by experts are classified incorrectly as background; True Negative (TN), indicating that the background pixels in the organoid images marked by experts are classified correctly as background; False Positive (FP), indicating that the background pixels in the images marked by experts are classified incorrectly as organoids. In order to evaluate the segmentation results, four evaluation indicators of accuracy (Acc, accuracy), sensitivity (Se, sensitivity), specificity (Sp, specificity), and F1-Score are used to evaluate the segmentation effect objectively. The definitions of evaluation indicators are shown in Table 3.

4.3 Results analysis of different segmentation models

This paper first tested the traditional segmentation algorithm to segment organoids. Three traditional segmentation algorithms were selected, which are the threshold segmentation algorithm (45), the canny edge detection algorithm (46), and the watershed algorithm (47). The results are shown in Figure 8. We can see from Figures 8C–E that the three traditional segmentation algorithms

TABLE 2 Experimental environment configuration.

Configuration	Parameter
OS	Ubuntu20.04
CPU	i9 7900
GPU	Geforce GTX 3090 GPU×2
CUDA	CUDA 11.1
Environment	Python 3.8
Frame	Pytorch 1.7.1

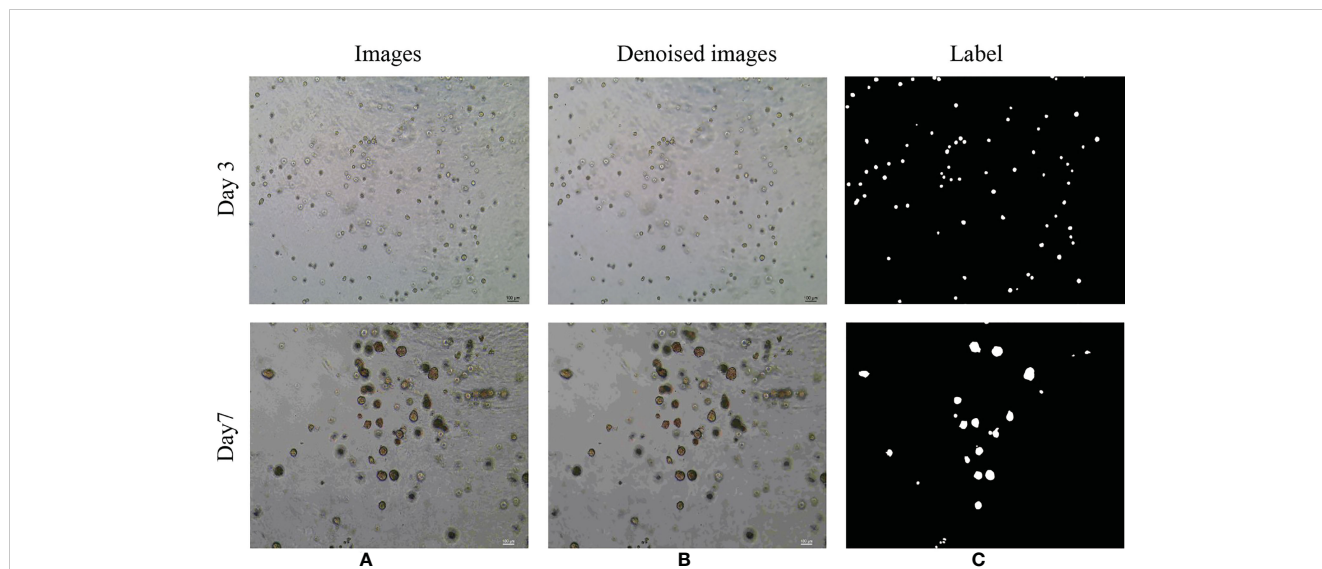


FIGURE 7 Examples of the original microscope image and corresponding denoised images of bladder cancer organoids. (A) Microscope images of bladder cancer organoids on culture day 3 and day 7 (B) Denoised images based on the microscope images of bladder cancer organoids. (C) The labels generated from denoised images of bladder cancer organoids.

have a good ability for organoid segmentation in a simple background, but it is quite different from tags for organoid segmentation in complex backgrounds. Next, we compared four algorithms of classical deep learning neural networks which are the FCN model, the U-Net model, the Attention U-Net model, the U2Net model, and our proposed ACU2Net model. These results are shown in Figures 8F–J. The segmentation results of the FCN model are not precise enough and incomplete, and there are a large number of unrecognized organoids. The segmentation results of the U-Net model have improved with obvious edges, but there are still problems of mis-segmentation in images with complex backgrounds. The Attention U-Net model has better segmentation results, and is less affected by background interference, but its segmentation outline is slightly rough. The U2Net model has precise segmentation edges and achieves better results in images with a large number of organoids, but in images with complex backgrounds, some organoids are often mis-segmented. In images with complex backgrounds and many organoids, the proposed ACU2Net network model can maintain a complete structure, and it can usually segment more detailed information in regions with weaker edges.

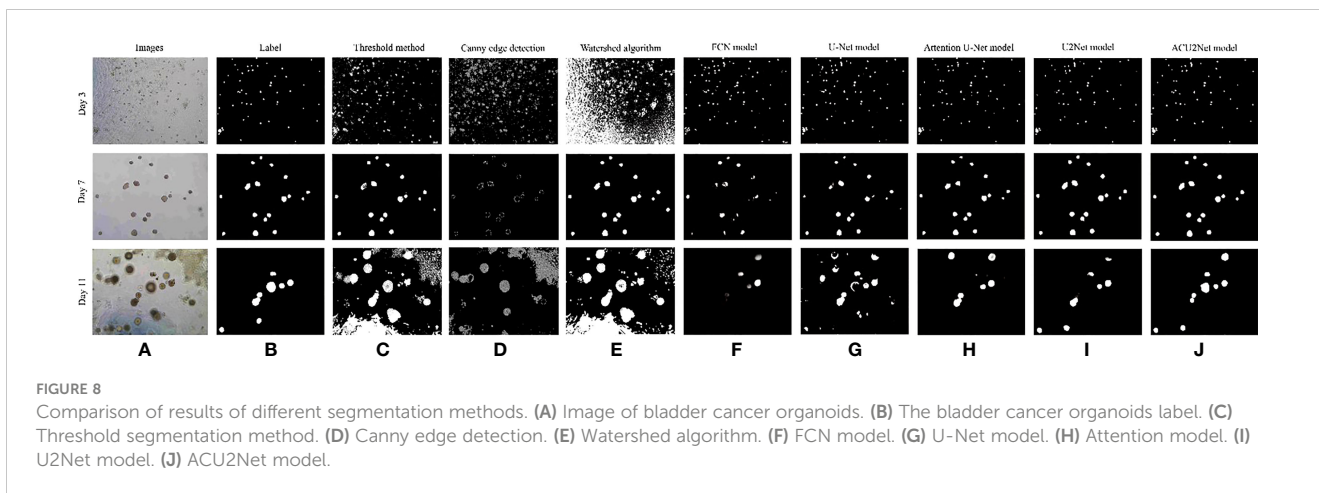
TABLE 3 The evaluation index.

Evaluation indicators	Formula
Acc	$Acc = \frac{TP + TN}{TP + TN + FP + FN}$
Se	$Se = \frac{TP}{TP + FN}$
Sp	$Sp = \frac{TN}{TN + FP}$
F1-Score	$F1 - Score = \frac{2 \times TP}{2 \times TP + FP + FN}$

Although it is more intuitive to observe with the naked eye, subjective factors may still be present, so it is still necessary to quantitatively evaluate the segmentation results. These results are shown in Table 4. The performance of our ACU2Net model is better than other models on the four evaluation indicators of Acc, Se, Sp, and F1-Score, which proves the rationality and effectiveness of the model. In addition, it can segment more organoid information in complex background regions, and has strong robustness. Due to the insufficient recovery of detailed information during the up-sampling process, the segmentation effect of the FCN model is slightly worse. By introducing deconvolution and feature layer connections, the U-Net model makes up for the lost detail information to a certain extent, and the effect is greatly improved. Furthermore, the Attention U-Net model uses the attention module to replace the traditional method of direct connection in the U-Net network, which enhances the effectiveness and selectivity of the connection in the effective connection of primary features and advanced features, and all indicators are significantly improved. The U2Net model utilizes a nested U-shaped network that integrates features of receptive fields of different sizes to segment most organoids. Compared with the U2Net model, the proposed model improved in various indicators, indicating that the proposed model can segment more organoids in complex areas and areas with blurred background edges, so it is more adaptable to interference factors such as contrast and noise, and the segmentation effect is better than other models.

4.4 The influence of each module on the overall model

In order to verify the effectiveness of the RAG and GCM modules introduced in this paper, the network was adjusted as follows (1): the original U2Net network, is represented by Net1 (2);



the U2Net network, which is combined with RAG module, is denoted by Net2; (3) the U2Net network, which is combined with GCM module, is represented by Net3; (4) The U2Net network, which is combined with RAG and GCM modules, is denoted by Net4. The influence of different modules on the model is shown in Table 5.

As can be seen from Table 4, the F1-score of the original U2Net network segmentation is 88.64. When the RAG module is added to the jump connection of the original U2Net network, the F1-Score of the network is 2.18% higher than that of the U2Net network, indicating that RAG attention can extract more detailed feature information about bladder cancer organoids and reduce the influence of a lot of background redundancy. It helps to combine high-level semantic information with low-level fine-grained surface information to improve the recognition ability. Compared with the U2Net network, the F1-Score of the GCM module is improved by 1.13%, indicating that the GCM module can maximize the feature transmission of significant regions of bladder cancer organoids, and the details of small targets would not be lost with the deepening of the feature

extraction process. Finally, when the two modules are added to the whole network simultaneously, the F1-Score is improved by 2.91%, indicating that the proposed algorithm has a good segmentation effect on the image of bladder cancer organoids with complex structures, which fully proves the effectiveness of the proposed algorithm. The test results of each module are shown in Figure 9.

The reasons for this improvement are: Based on the structure of U2Net, the channel attention network is inserted into the U2Net skip connection part, which improves the algorithm’s ability to learn the characteristics of bladder cancer organoids and the ability to process detailed information. As well,

ACU2Net can capture the spatial relationship with the surrounding pixels by applying high-resolution images in different layers of networks, but the ability to capture detailed features is poor. In addition, it can focus on the geometric details of the image by utilizing low-resolution images to obtain the local receptive field of the network, which is suitable for high-precision segmentation tasks, but the ability to represent images needs to be strengthened. Therefore, the different receptive fields obtained by the GCM module can further improve the

TABLE 4 Test results of different segmentation algorithms. Bold represents the best of each category.

Model	Acc(%)	Se(%)	Sp(%)	F1-Score(%)
FCN	97.83	83.47	81.40	82.42
U-Net	98.82	89.35	83.75	86.45
Attention U-Net	98.92	91.75	82.63	86.96
U2Net	98.97	92.64	84.97	88.64
ACU2Net	99.23	94.81	88.50	91.54

TABLE 5 Influence of each module on the whole model. Bold represents the best of each category.

	U2Net	RAG	GCM	Acc (%)	Se (%)	Sp (%)	F1-Score (%)
Net1	+			98.97	92.64	84.97	88.64
Net2	+	+		99.12	95.43	86.63	90.82
Net3	+		+	99.01	94.19	85.74	89.77
Net4	+	+	+	99.23	94.81	88.50	91.54

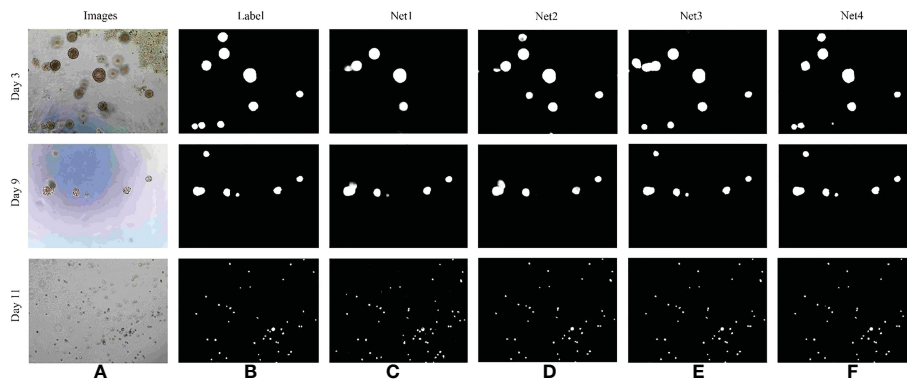


FIGURE 9 Comparison of results of different segmentation methods. (A) Image of bladder cancer organoids. (B) The bladder cancer organoids label. (C) Net1 model. (D) Net2 model. (E) Net3 model. (F) Net4 model.

image representation ability, to solve the problem of various sizes of bladder cancer organoids.

4.5 Drug screening evaluation

After training the model to identify bladder cancer organoids, segmentation is completed, and drug screening is required to select candidates for the treatment of bladder cancer for clinical trials. First, the segments as shown in Figure 10A need to be automatically computed to build a violin diagram, which can be drawn by calculating the organoid area after treatment with different drugs on different days. Figure 10B reflects the growth of bladder cancer organoids from 1 to 7 days in three environments, where the CTR group represents no drug treatment, and RA and B0107 are the abbreviations for the names of the two drugs used to treat bladder cancer, respectively. To observe the growth of organoids more intuitively, a quartile distribution map is added to the violin plot, and the dotted line in the middle is the median of this data group. As can be seen from the figure, as days 1-3 are the initial growth stage of bladder cancer organoids, there is no significant difference between

drug-treated and non-drug-treated organoids. The growth differences gradually appear after day 5, with the organoids in the CTR group reaching a peak. The median area of CTR is higher than those of RA and B0107, indicating that these two drugs have a specific inhibitory effect on the growth of bladder cancer organoids and can be put into clinical development. To verify the reliability of the experiment, we invited professionals in related fields to perform manual verification. These experts confirmed that the results of this paper showed the same regularity as the manual statistical results. While manual data processing time was 6 hours, the deep learning method used in this paper can screen and evaluate bladder cancer organoids in about 10 minutes. Therefore, the proposed method can promote the throughput for the testing of anti-cancer drugs for bladder cancer.

5 Conclusions

In this paper, an imaging segmentation method for bladder cancer organoids is proposed by using the U2Net basic framework combined with residual attention gate and grouping cross fusion module. The method employs GCM, the grouped cross merge module, to obtain

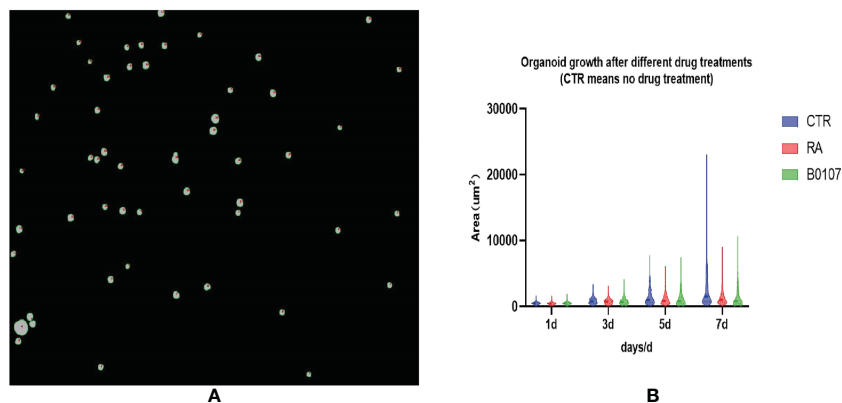


FIGURE 10 Drug screening evaluation based on ACU2Net model. (A) A micro image of organoids treated with drugs. (B) Using ACU2Net methods, the growth results of organoids after treatment of (retinoic acid) RA and B0107.

objects of different sizes at the skip connection of the model, which improves the feature representation of segmentation. Through integrating the improved attention mechanism, RAG, the semantic information of the feature is enhanced for finer segmentation, and the F1-Score indicator has reached 91.54% in this case. In addition, the application of this new method in evaluating the growth status of organoids on different culture days, with or without the treatment of drugs, found it could provide accurate screening results with higher efficiency for drug screening. This evidence shows that the proposed novel algorithm has a greater improvement in the imaging segmentation of bladder cancer organoids, especially in drug screening evaluation using bladder cancer organoids, than the existing algorithms. However, there are also some limitations to this method. Due to the complex background of the 2D images of bladder cancer organoids on Day 7, the adhesion of the organoids and the existence of aliasing, noise, and mixing effects of sampling points cannot be solved by image segmentation alone. In this case, the organoid volume for drug screening should also be considered. Therefore, the 3D reconstruction of bladder cancer organoids cultured for longer than seven days is the main focus of our later research.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

References

- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660
- Rheinwald JG, Green H. Serial cultivation of strains of human epidermal keratinocytes: the formation of keratinizing colonies from single cells. *Cell* (1975) 6(3):331–43. doi: 10.1016/S0092-8674(75)80001-8
- Lancaster MA, Knoblich JA. Organogenesis in a dish: Modeling development and disease using organoid technologies. *Science* (2014) 345(6194):1247125. doi: 10.1126/science.1247125
- Sato T, Vries RG, Snippert HJ, Wetering MD, Barker N, Daniel E, et al. Single Lgr5 stem cells build crypt-villus structures *in vitro* without a mesenchymal niche. *Nature* (2009) 459(7244):262–5. doi: 10.1038/nature07935
- Lce SH, Hu W, Matulay JT, Silva MV, Owczarek TB, Kim K, et al. Tumor evolution and drug response in patient-derived organoid models of bladder cancer. *Cell* (2018) 173(2):515–28. doi: 10.1016/j.cell.2018.03.017
- Mullenders J, de Jongh E, Brousalı A, Roosen M, Blom JPA, Begthelet H, et al. Mouse and human urothelial cancer organoids: A tool for bladder cancer research. *Proc Natl Acad Sci USA* (2019) 116(10):4567–74. doi: 10.1073/pnas.1803595116
- Kim E, Choi S, Kang B, Kong JH, Kim Y, Yoon WH, et al. Creation of bladder assembloids mimicking tissue regeneration and cancer. *Nature* (2020) 588:664–9. doi: 10.1038/s41586-020-3034-x
- Martinelli I, Tayebati SK, Tomassoni D, Nittari G, Roy P, Amenta F. Brain and retinal organoids for disease modeling: The importance of *In vitro* blood–brain and retinal barriers studies. *Cells* (2022) 11(7):1120. doi: 10.3390/cells11071120
- Betje J, Rindtorff N, Sauer J, Rauscher B, Dingert C, Gaitantzi H, et al. The drug-induced phenotypic landscape of colorectal cancer organoids. *Nat Commun* (2022) 13:3135. doi: 10.1038/s41467-022-30722-9
- Chen J, Fei N. Organoid technology and applications in lung diseases: Models, mechanism research and therapy opportunities. *Front Bioeng Biotechnol* (2022) 10:1066869. doi: 10.3389/fbioe.2022.1066869
- Gong Z, Xu H, Su Y, Wu WF, Hao L, Han CH. Establishment of a Novel Bladder cancer xenograft model in humanized immu-nodeficient mice. *Cell Physiol Biochem* (2015) 37(4):1355–68. doi: 10.1159/000430401
- Elbadawy M, Usui T, Mori T, Tsunedomi R, Hazama S, Nabeta R, et al. Establishment of a novel experimental model for muscle-invasive bladder cancer using a dog bladder cancer organoid culture. *Cancer Sci* (2019) 110(9):2806–21. doi: 10.1111/cas.14118
- Molefi T, Marima R, Demetriou D, Afra B, Zodwa D. Employing AI-powered decision support systems in recommending the most effective therapeutic approaches for individual cancer patients: Maximising therapeutic efficacy. (2023). doi: 10.1007/978-3-031-21506-3_13
- bugomaa A, Elbadawy M. Patient-derived organoid analysis of drug resistance in precision medicine: is there a value? *Expert Rev Precis Med Drug Dev* (2020) 5(1):1–5. doi: 10.1080/23808993.2020.1715794

Funding

This paper is supported by the National Natural Science Foundation of China (62066046); National Natural Science Foundation of China, No.32070818; Key Project of Science and Technology Department of Yunnan Province, No.202001B050005; the National Natural Science Foundation of China (Grant No. 82204695); the Sichuan Provincial Research Institutes Basic Research Operations Fund Project (Grant No. A-2022N-Z-2); and the Sichuan Academy of Traditional Chinese Medicine Research Project (Grant No. QNCJRSC2022-9).

Acknowledgments

We wish to thank all of the participants included in this study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

15. Zhang S, Yu P, Li H, Li HS. (2022). Research on retinal vessel segmentation algorithm based on deep learning, in: *2022 IEEE 6th Information Technology and Mechatronics Engineering Conference (ITOEC)*. IEEE, Vol. 6. pp. 443–8. doi: 10.1109/ITOEC53115.2022.9734539
16. Setio AAA, Ciompi F, Litjens G, Gerke P, Jacobs C, Riel SJV, et al. Pulmonary nodule detection in CT images: False positive reduction using multi-view convolutional networks. *IEEE Trans Med Imaging* (2016) 35(5):1160–9. doi: 10.1109/TMI.2016.2536809
17. Qin XB, Zhang Z, Huang C, Dehghan M, Zaiane OR, Jagersand M, et al. U2-net: Going deeper with nested U-structure for salient object detection. *Pattern Recognit* (2020) 106:107404. doi: 10.1016/j.patcog.2020.107404
18. Mnih V, Hees N, Gravs A. Recurrent models of visual attention. *Adv Neural Inf Process Syst* (2014) 3:2204–12. doi: 10.48550/arXiv.1406.6247
19. Valler MJ, Green D. Diversity screening versus focused screening in drug discovery. *Drug Discovery Today* (2000) 5(7):286–93. doi: 10.1016/S1359-6446(00)01517-8
20. Argenio DZD, Park K. Uncertain pharmacokinetic/pharmacodynamic systems: design, estimation and control. *Control Eng Pract* (1997) 5(7):1707–16. doi: 10.1016/S0967-0661(97)10025-9
21. Holford N, Ma SC, Ploeger BC. Clinical trial simulation: a review. *Clin Pharmacol Ther* (2010) 88(2):166–82. doi: 10.1038/clpt.2010.114
22. Schneider G, Fechner U. Computer-based *de novo* design of drug-like molecules. *Nat Rev Drug Discovery* (2005) 4(8):649–63. doi: 10.1038/nrd1799
23. Hinton GE, Osindero S, Teh. YW. A fast learning algorithm for deep belief nets. *Neural Comput* (2006) 18(7):1527–54. doi: 10.1162/neco.2006.18.7.1527
24. Irwin JJ, Shoichet BK, Mysinger MM, Huang N, Colizzi F, Wassam P, et al. Automated docking screens: a feasibility study. *J Med Chem* (2009) 52(18):5712–20. doi: 10.1021/jm9006966
25. Tsai TY, Chang KW, Chen. CYC. IScreen: world's first cloud-computing web server for virtual screening and *de novo* drug design based on CM database at Taiwan. *J Comput-Aided Mol Design* (2011) 25(6):525–31. doi: 10.1007/s10822-011-9438-9
26. Zhu Y, Huang R, Zhu R, Xu W, Zhu RR, Cheng L. DeepScreen: An accurate, rapid, and anti-interference screening approach for nanoformulated medication by deep learning. *Adv Sci* (2018) 5(9). doi: 10.1002/advs.201800909
27. Yu KH, Beam AL, Kohane. IS. Artificial intelligence in healthcare. *Nat Biomed Eng* (2018) 2(10):719–31. doi: 10.1038/s41551-018-0305-z
28. Li S, Song W, Fang L, Ghamisi P, Benediktsson JA. Deep learning for hyperspectral image classification: An overview. *IEEE Trans Geosci Remote Sens* (2019) 57(9):6690–709. doi: 10.1007/978-981-33-4420-4_6
29. Ghosh S, Das N, Das I, Maulik U. Understanding deep learning techniques for image segmentation. *ACM Comput Surveys (CSUR)* (2019) 52(4):1–35. doi: 10.1145/3329784
30. Zhao ZQ, Zheng P, Xu S, Wu X. Object detection with deep learning: A review. *IEEE Trans Neural Networks Learn Syst* (2019) 30(11):3212–32. doi: 10.1109/TNNLS.2018.2876865
31. Shelhamer E, Long J, Darrell. T. Fully convolutional networks for semantic segmentation. *IEEE Trans Pattern Anal Mach Intell* (2017) 39(4):3431–40. doi: 10.1109/CVPR.2015.7298965
32. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. *Springer* (2015) 234–41. doi: 10.1007/978-3-319-24574-4_28
33. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang M. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Trans Med Imaging* (2019) 39(6):1856–186. doi: 10.1109/TMI.2019.2959609
34. Olimov B, Sanjar K, Din S, Anand A, Paul A, Kim JH. FU-net: fast biomedical image segmentation model based on bottleneck convolution layers. *Multimedia Syst* (2021) 27(4):637–50. doi: 10.1007/s00530-020-00726-w
35. Quan TM, Hildebrand DGC, Jeong. WK. Fusionnet: A deep fully residual convolutional neural network for image segmentation in connectomics. *Front Comput Sci* (2021) 3. doi: 10.3389/fcomp.2021.613981
36. Fu J, Liu J, Tian H, Dai J. (2019). Dual attention network for scene segmentation, in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 3146–54. doi: 10.1109/cvpr.2019.00326
37. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, et al. Attention u-net: learning where to look for the pancreas. *arXiv preprint arXiv* (2018) 2019:1903.11834. doi: 10.48550/arXiv.1804.03999
38. Li Z, Wang SH, Fan RR, Cao G, Zhang YD, Guo T. Teeth category classification via seven-layer deep convolutional neural network with max pooling and global average pooling. *Int J Imaging Syst Technol* (2019) 29(4):577–83. doi: 10.1002/ima.22337
39. Liu K, Kang G, Zhang N, Hou B. Breast cancer classification based on fully-connected layer first convolutional neural networks. *IEEE Access* (2018) 6:23722–32. doi: 10.1109/ACCESS.2018.2817593
40. Pieter TDB, Kroese DP, Mannoret S, Rubinstein RY. A tutorial on the cross-entropy method. *Ann OR* (2005) 134(1):19–67. doi: 10.1007/s10479-005-5724-z
41. Shen C, Roth HR, Oda H, Oda M, Hayashi Y, Misawa K, et al. *On the influence of dice loss function in multi-class organ segmentation of abdominal CT using 3D fully convolutional network*. (2017) 281:117. doi: 10.48550/arXiv.1801.05912
42. Buades A, Coll B, Morel. JM. (2005). A non-local algorithm for image denoising, in: *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2. pp. 60–5. doi: 10.1109/CVPR.2005.38
43. Zeng K, Liu J, Jiang Z, Xu D. A decreasing scaling transition scheme from Adam to SGD. *Adv Theory Simulations* (2022) 2100599. doi: 10.48550/arXiv.2106.06749
44. Branco T, Staras K. The probability of neurotransmitter release: variability and feedback control at single synapses. *Nat Rev Neurosci* (2009) 10:373–83. doi: 10.1038/nrn2634
45. Kohler R. A segmentation system based on thresholding. *Comput Graphics Image Process* (1981) 15(4):319–38. doi: 10.1016/S0146-664X(81)80015-9
46. Bao P, Zhang L, Wu. X. Canny edge detection enhancement by scale multiplication. *IEEE Trans Pattern Anal Mach Intell* (2005) 27(9):1485–90. doi: 10.1109/TPAMI.2005.173
47. Biemiek A, Moga A. An efficient watershed algorithm based on connected components. *Pattern recognition* (2000) 33(6):907–16. doi: 10.1016/S0031-3203(99)00154-5