

Article

GCT-UNET: U-Net Image Segmentation Model for a Small Sample of Adherent Bone Marrow Cells Based on a Gated Channel Transform Module

Jing Qin ¹, Tong Liu ¹, Zumin Wang ^{2,*}, Lu Liu ^{3,*} and Hui Fang ⁴¹ School of Software Engineering, Dalian University, Dalian 116622, China² College of Information Engineering, Dalian University, Dalian 116622, China³ Heart Centre, Affiliated Zhongshan Hospital of Dalian University, Dalian 116023, China⁴ Department of Computer Science, Loughborough University, Loughborough LE11 3TU, UK

* Correspondence: zumingwang@163.com or wangzumin@dlu.edu.cn (Z.W.); liulu2162192@163.com or 101279@yzpc.edu.cn (L.L.)



Citation: Qin, J.; Liu, T.; Wang, Z.; Liu, L.; Fang, H. GCT-UNET: U-Net Image Segmentation Model for a Small Sample of Adherent Bone Marrow Cells Based on a Gated Channel Transform Module. *Electronics* **2022**, *11*, 3755. <https://doi.org/10.3390/electronics11223755>

Academic Editor: Luigi Landini

Received: 29 September 2022

Accepted: 14 November 2022

Published: 16 November 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/>).

Abstract: Pathological diagnosis is considered to be declarative and authoritative. However, reading pathology slides is a challenging task. Different parts of the section are taken and read for different purposes and with different focuses, which further adds difficulty to the pathologist's diagnosis. In recent years, the deep neural network has made great progress in the direction of computer vision and the main approach to image segmentation is the use of convolutional neural networks, through which the spatial properties of the data are captured. Among a wide variety of different network structures, one of the more representative ones is UNET with encoder and decoder structures. The biggest advantage of traditional UNET is that it can still perform well with a small number of samples, but because the information in the feature map is lost in the downsampling process of UNET, and a large amount of spatially accurate detailed information is lost in the decoding part. This makes it difficult to complete accurate segmentation of cell images with dense numbers and high adhesion. For this reason, we propose a new network structure based on UNET, which can be used to segment cell images by aggregating the global contextual information between different channels and assigning different weights to the corresponding channels through the gated adaptive mechanism, we improve the performance of UNET in the cell segmentation task and consider the use of unsupervised segmentation methods for secondary segmentation of the predicted results of our model, and the final results obtained are tested to meet the needs of the readers.

Keywords: bone marrow cell segmentation; bone marrow cell count; UNET; gated channel transformation

1. Introduction

Pathological diagnosis is the most reliable method of disease examination and is even regarded as the gold standard for tumor diagnosis. The final report of the pathology section requires the high skill and experience of pathologists. The purpose and focus of reading the section vary depending on the site of sampling, which further increases the difficulty of reading the section. Different pathologists may obtain different results, and statistically, physician misclassification of myeloid cells during microscopic examination is associated with misclassification of the number of myeloid cells [1]. This technique was developed based on HE-stained pathological sections of radiation-injured rat femurs as a study medium, to contribute to a more objective statistical analysis of bone marrow cell changes by relevant investigators, significantly improving test efficacy and saving investigators' time. In this regard, it is urgent to improve the accuracy of counting bone marrow cells during microscopy, and one of the directions of its solution is the use of computerized microscopy systems [2,3]. Automatic cell morphology detection systems based on computer

vision techniques have been widely used in recent years. This system aids physicians in diagnosis mainly by simulating a manual microscope, which achieves a high degree of objectivity in the obtained results and greatly reduces the work intensity of physicians, further improving the efficiency and diagnostic accuracy of pathological testing. Whether the bone marrow cell image can be accurately segmented greatly affects the accuracy of the classification and counting of different bone marrow cell types. Therefore, accurate cell segmentation of bone marrow pathology sections is an important and challenging task due to the characteristics of bone marrow cells and the uncertainty of available system imaging [4]. For image segmentation, to obtain automatic segmentation of regions of interest, many researchers have investigated many image segmentation methods based on unsupervised and supervised learning algorithms.

The following image segmentation methods are commonly used in the unsupervised domain, threshold-based segmentation methods [5], which usually separate the selected background from the foreground by setting a specific threshold value. Region-based segmentation methods [6], a mathematical morphological-based segmentation approach [7]. The algorithm is compared to the threshold and region growth methods, it can produce stable segmentation results and has been widely used in the field of medical image segmentation. Traditional classification algorithms such as Support Vector Machines (SVM) and plain Bayesian classifiers have also been used as classifiers for cells [8,9]. Ref. [10] proposed a region-based method, which approximates any 2D shape by automatically determining the number of possible overlapping ellipses. Although these segmentation methods have achieved good performance, they still cannot be applied to the automatic segmentation of bone marrow cell smears. This is because the traditional methods have the limitations of low learning ability and the need for manual features. Moreover, due to the large number of cells and the high degree of adhesion in bone marrow pathological sections, it is impossible to separate the boundaries of cell adhesions smoothly using these methods, and thus, the counting of bone marrow cells cannot be completed.

With the continuous development in the field of deep learning, convolutional neural networks with representational learning capability have shown excellent performance in the field of image recognition and classification, and have been successfully applied to image classification [11], object detection [12] and image segmentation [13] fields. Some neural network structures have been applied to image segmentation. For example, the Full Convolutional Network (FCN) is a network structure consisting of only convolutional layers [13]. The biggest feature of the FCN network is that it has no requirements for the size of the original image input, that is, it can accept any size of image input, which can significantly reduce the workload of preprocessing. DeepLab [14], based on multi-scale cavity convolution, adopts conditional random field to improve the boundary segmentation effect of the model. The encoder-decoder architecture is also used for semantic segmentation. It includes an encoder using a convolution layer and pooling layer to extract image features and a decoder combining the extracted features for final classification, such as UNET [15]. A variant network UNET++ [16] by nesting multiple UNET considers and optimizes the traditional UNET with some problems and obtained good performance. The attention-based UNET [17], the segmentation performance of UNET is improved by introducing an attention mechanism so that the values of irrelevant regions become smaller and the values of target regions become larger. GRUU-UNET [18], which combines both CNN multiscale features and RNN iterative refinement models, UNET constructed based on the transformer module [19] both obtained better segmentation results than the traditional CNN-UNET. In terms of method fusion, a cell segmentation model combines marker-controlled watershed transform and deep learning methods [20]. For the problem of bone marrow cell identification and segmentation, there are also many related studies experimenting with deep learning methods, such as [21], which proposed to identify and segment CD138+ and CD138- stained cells in bone marrow cells using a semantic segmentation-based convolutional neural network. The authors in [4] proposed to use a YOLOv5 network-based bone marrow cell detection algorithm, trained by minimizing a novel loss function, and the

results show that the proposed loss function effectively improves the performance of the algorithm. Ref. [22] proposed a color space transformation and multi-class weighted loss scheme to optimize the UNET in white blood cell segmentation performance. However, although researchers have achieved basic segmentation of bone marrow cells, the field of how to separate densely populated and highly adherent bone marrow cells for automatic counting of bone marrow cells has not been investigated. Therefore, this paper focuses on a deep learning-based algorithm for bone marrow cell detection, which targets the dense distribution of cells in bone marrow sections, the large number of adherent cells, and the difficulty of segmentation due to the large degree of adhesion. In this paper, we design a new structure GCT (Gated Channel Transformation)-UNET based on the traditional UNET network structure, which improves the performance of UNET in cell segmentation tasks by focusing on the differences between image channels and space through the gated channel transformation module. At the same time, compared with the traditional UNET network, our network does not introduce too many additional parameters, so the training speed of the model is also better, and the performance is also better on the data set with a small number of samples, which is very meaningful for the problem of difficult acquisition of medical image data and a small amount of data. The main contributions of this paper are as follows:

- A UNET neural network model combined with a gated transformation module is proposed. In the case of a small amount of data, the characteristics of UNET with few parameters and fast training speed are preserved as much as possible, and the cell segmentation effect is improved at the same time. In particular, our proposed model performs better in segmenting adherent cells in microscopic images of bone marrow cells compared to some currently existing neural network models.
- We propose a secondary segmentation idea for high-adhesion areas of bone marrow cells, that is, using the watershed algorithm to post-process the predictions of the model to obtain predictions that are closer to the label.

2. Materials and Methods

In this study, we propose a GCT-UNET model that aims to be able to segment more accurately the adherent, dense cells in the cellular images of bone marrow pathology sections. The overall process we designed is shown in Figure 1. In our proposed method, the first step is to preprocess the original data, mainly adjusting the size of the original image so that it can adapt to the following operations, then making the corresponding mask label map for the appropriate size image, and applying the data enhancement method to the image and the corresponding mask label map. Finally, these pictures were sent to the GCT-UNET network we developed for training. At the beginning, the accuracy of the prediction results obtained by the network was very low, and the effect was very poor. Therefore, we optimized the parameters in the network through backpropagation. After repeating the above steps many times, we finally obtained a robust model that met our expectations. Finally, we implement an automatic system that can segment the pathological section images of bone marrow cells quickly.

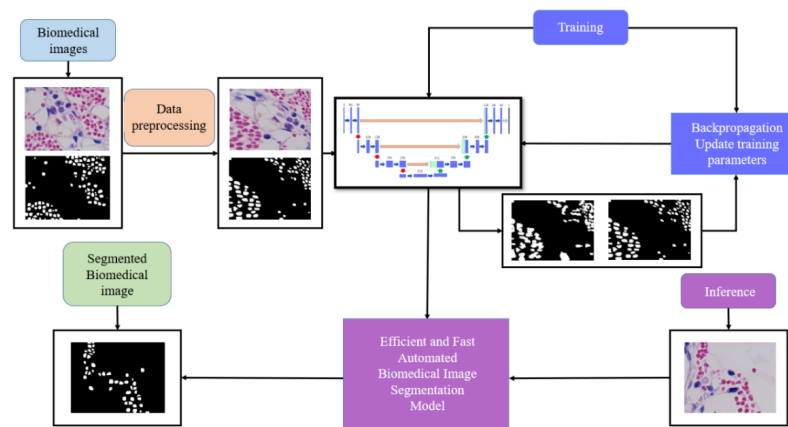


Figure 1. The proposed methodology.

2.1. Network Architecture

Our structure follows the same U-shaped structure as UNET, as shown in Figure 2. The encoder and decoder parts are also included. Compared with the traditional UNET five-layer convolutional module, we consider only a four-layer convolutional module in order to fit the segmentation task in this paper and to avoid the excessive number of parameters. For the encoder part, each convolutional block contains two 3×3 sized convolutional layers, and to keep the feature map size constant before and after convolution, we use edge filling for each convolutional layer. After each convolutional layer, we add a gated channel transform module, a modified linear unit (ReLU) and a batch normalization function. Each convolutional block is followed by a maximum pooling layer for downsampling to reduce network complexity and compress the features. After encoding the image data in the network, the model has learned the important features of the image, but is not ready to generate the segmented image at this time, because although the model knows some features contained in the image, it still does not have enough information about the location of these features and therefore needs to decode them. For the decoder part, each convolutional block contains the same content inside as the decoder part. After each convolutional block, a bilinear interpolation upsampling is performed twice to increase the image size, and then a jump connection is made to the corresponding output feature map of the encoder part. After the decoding operation, we use a 1×1 convolutional layer after the last set of decoder blocks, which is used to generate the final segmented image for our model.

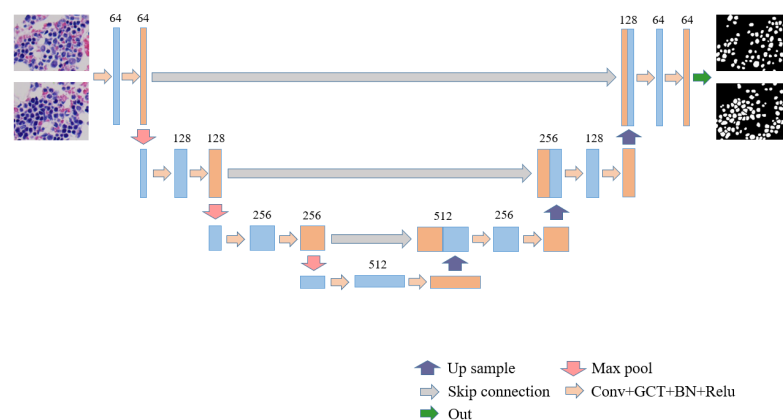


Figure 2. The structure of our proposed model.

2.2. Convolution Blocks

The main purpose of convolution layers (2D convolution layers) is to extract features from an image. By superimposing multiple convolution layers, we can obtain deeper

abstract features of the original image, such as detecting edges in the first layer, then simpler shapes in the second layer, and then higher-level features. Moreover, any unit in the output image is related to only a part of the input image so that each region has its own exclusive features and is not influenced by other regions [23]. The rectified linear unit (ReLU) is used as the activation function of the neuron. The ReLU activation function can be expressed by the following equation.

$$ReLU = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases} \quad (1)$$

The main role of pooling layers is to reduce the dimensionality of the feature map and reduce the complexity of the neural network. The maximum pooling is mainly used to reduce the second error, so we choose the maximum pooling for downsampling, which can preserve more texture information of the image.

2.3. Gated Channel Transformation Module

Gated Channel Transformation [24] is a module that normalizes channels without introducing parameters. It can determine whether to activate or inhibit channel expression by capturing the relationship between channels. In order to adapt the training process of the neural network, it also assigns weights to each channel, so that the input characteristics can be automatically adjusted during the network training process. It can be expressed by the following formula:

$$\hat{X} = F(X|\alpha, \gamma, \beta), \alpha, \gamma, \beta \in R^C \quad (2)$$

C represents the number of channels of the feature map, α , γ and β It is the parameters that can be trained with the network, and together they determine the behavior in each GCT module.

3. Experiments and Results

3.1. Dataset

HE-stained pathological sections of the femur of radiation-injured SD rats in medical research. The bone marrow changed significantly before and after radiation, and the blue-stained cells (hereafter referred to as blue cells) in the field of view were replaced by red-stained cells (hereafter referred to as red cells) to varying degrees depending on the group. The real microscopic image data of bone marrow cells acquired by this technique contains both real images and masked images made under the guidance of a physician. In total, the original images contain 15 microscopic images in tif format of 2748×2200 pixels with blue cells dominating and 19 microscopic images in tif format of 2748×2200 pixels with red cells dominating. In order to reduce the workload when making masks, we selected microscopic images with a small number of cells for masking as much as possible without losing information.

3.2. Data Pre-Processing

Through the analysis, we decided to segment the blue cells and red cells separately. Firstly, the original images were preprocessed for segmentation, where 15 blue original cell images were segmented into 240 images of 340×256 pixels in jpg format and 19 red original cell images were segmented into 304 images of 340×256 pixels in jpg format. Then, 53 images were selected from this segmented blue image group and 52 images were selected from the segmented red image group, and image masks were made for them accordingly. For the blue image group, we apply data enhancement techniques, including rotation, flipping, and perspective shifting, to both the 53 original images and the 53 masked images. The blue image set was expanded to contain 162 samples and used as our blue cell training data set, and the same method was used to expand the 52 red images set to contain 200 samples and used as our red cell training data set.

3.3. Experimental Details

The model is deployed on a host with dual AMD CPUs and two GTX1080ti graphics cards, and the deep learning model is implemented using the Pytorch platform and Python 3.7. Image pre-processing and final counting are implemented by the Matlab platform.

3.4. Loss Function

In this paper, a joint loss function combining dice loss [25] and binary cross-entropy loss is used. Dice loss is derived from the dice coefficient, a measurement function which is used to evaluate the similarity between samples. The value is in the $[0, 1]$ interval, the larger the value the more similar the two samples are. The dice coefficient is defined as follows:

$$\text{dice} = \frac{2|X \cap Y|}{|X| + |Y|} \quad (3)$$

where $|X \cap Y|$ is the intersection between X and Y , $|X|$ and $|Y|$ represent the number of elements of X and Y , and the numerator is multiplied by 2 in order to ensure that the denominator takes values in the range $[0, 1]$ after repeated calculations. The dichotomous problem dice coefficients can also be expressed as:

$$\text{dice} = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (4)$$

Then, dice loss can be expressed as follows:

$$\text{dice loss} = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \quad (5)$$

The binary cross-entropy loss applies to the binary classification task. Since the binary classification task has only positive and negative cases, and the probability sum of both is 1, the binary cross-entropy loss function can be defined in a simplified form as follows:

$$\text{bceloss} = -(y \cdot \log(\hat{y}) + (1 - y) \cdot \log(1 - \hat{y})) \quad (6)$$

where y is the labeled sample, which takes the value of 1 if the sample belongs to the positive example and 0 otherwise. Thus, the joint loss function used in this paper can then be expressed as:

$$\text{dicebceloss} = \text{bceloss} * \omega + \text{diceloss} * (1 - \omega) \quad (7)$$

where ω is the weighting factor, which often takes the value of 0.5.

3.5. Evaluation Indicators

In order to validate the effectiveness and robustness of our proposed model, we evaluated the considered model against several evaluation metrics applicable to the image segmentation task. Since the objective considered in this paper is for the segmentation of densely adherent cells, the ratio of foreground to background in the employed data images is close to 1. Therefore, the first evaluation metric we use is the Intersection over Union (IoU) ratio, which returns a value equal to the overlap region divided by the union region between the prediction mask and the target mask. Based on the confusion matrix it can be expressed by the following equation:

$$\text{IoU} = \frac{TP}{TP + FP + FN} \quad (8)$$

Pixel Accuracy (PA) is another metric to evaluate the considered model and it represents the percentage of pixels predicted by the model to the total pixels. Based on the confusion matrix it can be expressed by the following equation:

$$PA = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

where TP, FP, TN, FN are denoted as true case, false positive case, true negative case and false negative case.

3.6. Result

After completing the training and validation of the model, we tested it on a test set, which was divided into a blue cell test set and a red cell test set with 15 and 16 images of 340×256 sizes, respectively. Based on the evaluation metrics described in Section 3.5 we evaluated U-NET, SE-UNET [26], Attention-UNET, and GCT-UNET (the comparison results are shown in Table 1), and Figure 3 shows the trends of training and validation losses corresponding to the different models. To ensure the test persuasiveness, we set all the objects to be tested to have the same number of layers.

Table 1. Comparison of experimental results of different models.

Methodology	IOU		PA		Parameter Quantity (Mb)
	Blue Cell	Red Cell	Blue Cell	Red Cell	
KMeans	0.479	0.414	0.727	0.641	
Fuzzy C-Means	0.448	0.404	0.688	0.610	
UNET	0.764	0.727	0.909	0.848	7.788
Attention-UNET	0.771	0.731	0.918	0.855	8.651
SE-UNET	0.769	0.733	0.914	0.859	7.843
GCT-UNET	0.776	0.739	0.921	0.863	7.796

From Table 1, we can see that our proposed model performs slightly better for a small number of samples and improves the expression of important features by controlling the relationship between the channels. At the same time, compared with the traditional attention mechanism, since GCT does not introduce additional convolutional layers to the model, it can improve the model performance while preserving as much as possible the small number of UNET parameters and the fast computing speed fast.

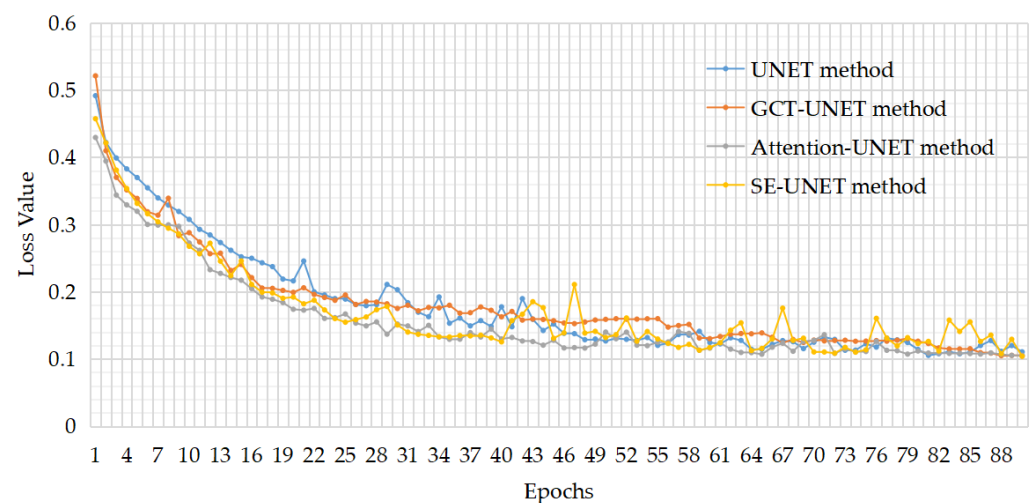


Figure 3. Validation set loss decline curves for different model training.

It can be observed from Figure 3 that when the number of iterations is the same, the loss value of our model fluctuates less during the decline process, which indicates that

the model is more stable during the convergence process. Figure 4 shows the original images and labels of some datasets, the prediction results of our model and the prediction results of some other models. It can be observed in Figure 4 that our proposed model performs slightly better than UNET, Attention-UNET and SE-UNET for dense adherent cell segmentation with few data samples, especially when the cell morphology is irregular and the adhesion degree is large, the noise is less than other methods, it is more helpful to achieve the cell counting requirement. The main reason we believe this is that the gated transition module tends to encourage more cooperation at the shallow level, while at the deeper level it tends to promote competition. The shallow level mainly captures general texture features, this shallow texture feature is less useful for the segmentation of adherent cells, while the deeper high-level features have stronger discriminative power and task relevance, so this part of features is more significant for segmenting adherent cells. Since the boundary of the adherent part of the cell is blurred, the model is better made to notice the change of the boundary pixels by strengthening the competition between the channels in this part during training, and combining the features of the neighboring pixels of the boundary pixels to obtain better segmentation of the adherent cells while for the detailed parts that cannot be separated by our proposed model, post-processing using the watershed algorithm [27] can be considered to obtain further optimization of the segmentation effect.

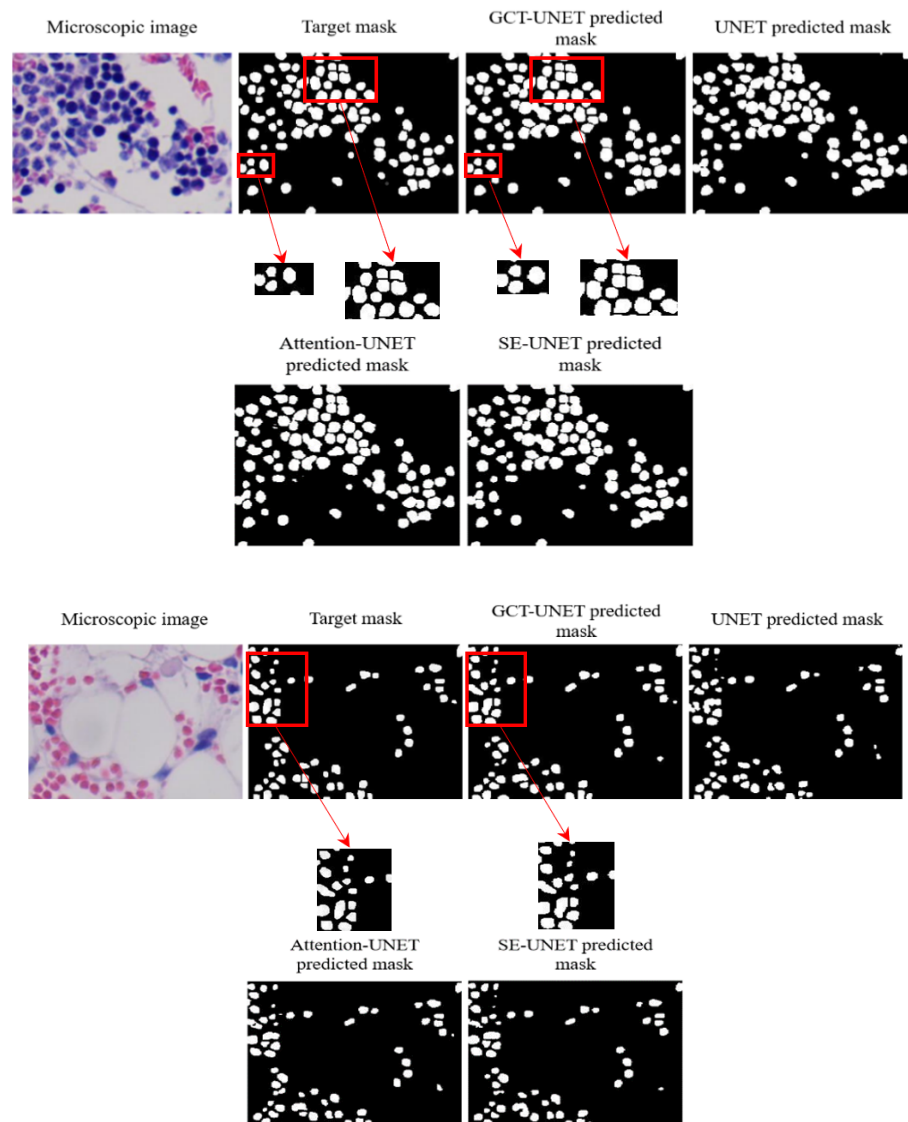


Figure 4. Compare the segmentation results of different models.

Figure 5 shows the post-processing results of an example in Figure 4. As can be seen from Figure 5, since the watershed algorithm is an unsupervised algorithm, it mainly segments the image according to the similarity between pixels and the corresponding segmentation criteria. Since our proposed model predicts less noise in the segmented images compared to other methods, although there may be a small amount of over-segmentation after post-processing using the watershed method, but for the parts that our model cannot segment, after comparison we use watershed post-processing of the algorithm will give better results.

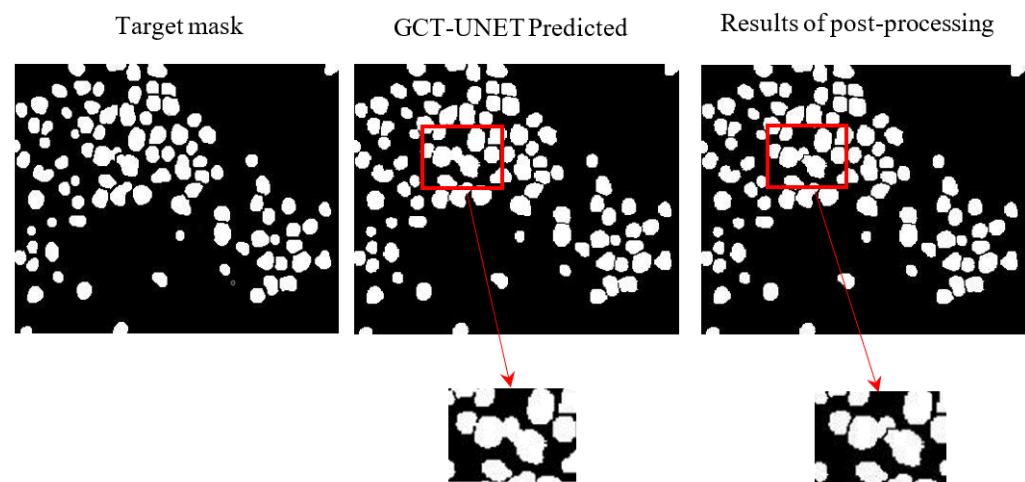


Figure 5. Compare the results of our model generation with the results of secondary segmentation.

Figure 6 shows the results that are not very effective in our work. It can be seen that in the case of excessive cell density and irregular cell morphology, although the method proposed by us will generate a small amount of excessive segmentation and false segmentation, resulting in an inaccurate final count, we can generally accept a small number of such situations in our diagnostic tasks.

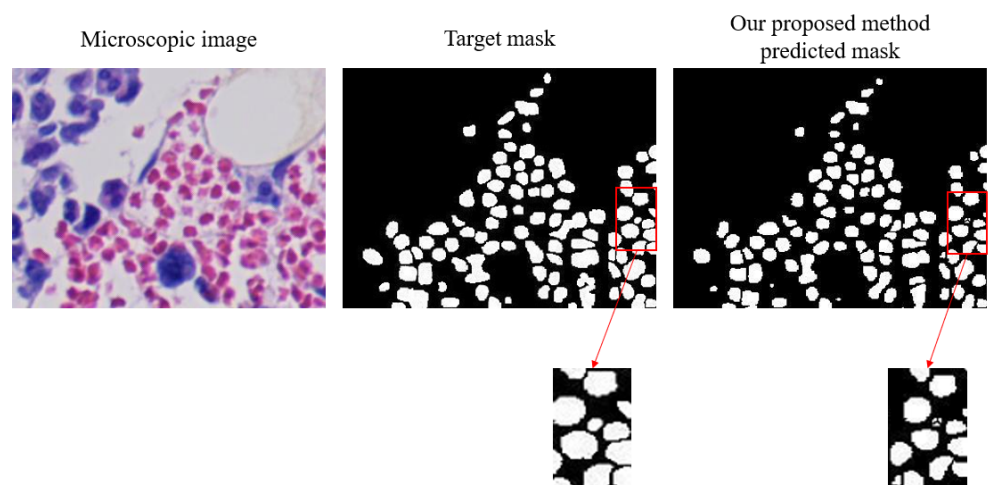


Figure 6. Examples of poor results.

4. Conclusions

In this paper, we propose the GCT-UNET model, which aims to improve the performance of the existing CNN-based biomedical image segmentation model U-Net. The conventional UNET is not effective in segmenting heavily adhered and dense cells, and it may cause segmentation errors and incomplete segmentation of adhered cells, so it cannot fulfill the demand for automatic cell counting. By introducing the gated channel transformation module to the traditional UNET, we can make the model more sensitive to the pixel change

of the adherent part of the cells by exploiting the characterization ability of each channel and assigning different weights to it, so that the model can segment the adherent cells more accurately. Moreover, compared with some other UNET variants, such as Attention-UNET and SE-UNET mentioned in Table 1, our model seems to be more stable in terms of convergence during training while improving the segmentation effect. In addition, we do not use the dropout regularization scheme, but use the regularization ability of batch normalization technology, which can accelerate the training speed of network models, which is helpful for data sets with small data volumes. During the training, we observed that there was no problem fitting the model. We tested the segmentation time by loading the model parameters for a single image of 340×256 sizes with UNET, Attention-UNET, SE-UNET, and GCT-UNET corresponding to segmentation times of 2.87 s, 3.33 s, 2.99 s, and 2.96 s. The comparison revealed that our model improved the stability and objectivity of counting results while saving the time cost of researchers, and its application in analyzing images of radiation-injured bone marrow cell changes is currently in use.

In the future, we hope to obtain better segmentation results by optimizing the model parameters and applying them to future better base models, as well as to optimize and apply the model to different types of medical data with different segmentation targets, to improve the effectiveness of computer intelligence in segmenting medical images and to assist doctors in better-performing related tasks.

Author Contributions: Methodology, J.Q. and H.F.; Investigation, L.L.; Data curation, T.L.; Supervision, Z.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Youth Fund Project of the National Nature Fund of China under Grant 62002038. This work was supported by the Hexi University scientific research innovation and application Xiaozhang Fund key projects XZZD2019006.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of Animal Experimental Ethical Inspection of Gansu University of Chinese Medicine (protocol code 2022-086, 3 March 2022).

Informed Consent Statement: Not applicable. The study is just involved HE-stained pathological sections of the femur of radiation-injured SD rats in medical research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kazemi, F.; Najafabadi, T.A.; Araabi, B.N. Automatic recognition of acute myelogenous leukemia in blood microscopic images using k-means clustering and support vector machine. *J. Med. Signals Sens.* **2016**, *6*, 183. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Nikitaev, V. High-tech information-measuring complexes of oncological diagnostics: problems and key provisions of the construction methodology. *Izmer. Tekhn* **2015**, 68–70.
3. Nikitaev, V. Modern principles of measurements in intelligent systems of histological diagnostics of oncological diseases. *Izmer. Tekhn* **2015**, 68–70.
4. Huang, D.; Cheng, J.; Fan, R.; Su, Z.; Ma, Q.; Li, J. Bone marrow cell recognition: Training deep object detection with a new loss function. In Proceedings of the 2021 IEEE International Conference on Imaging Systems and Techniques (IST), San Francisco, CA, USA, 23–26 October 2021; pp. 1–6.
5. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [\[CrossRef\]](#)
6. Adams, R.; Bischof, L. Seeded region growing. *IEEE Trans. Pattern Anal. Mach. Intell.* **1994**, *16*, 641–647. [\[CrossRef\]](#)
7. Beucher, S.; Meyer, F. The morphological approach to segmentation: The watershed transformation. In *Mathematical Morphology in Image Processing*; CRC Press: Hoboken, NJ, USA, 2018; pp. 433–481.
8. Prinyakupt, J.; Pluempitiwiriawej, C. Segmentation of white blood cells and comparison of cell morphology by linear and naïve Bayes classifiers. *Biomed. Eng. Online* **2015**, *14*, 63. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Rezaatofghi, S.H.; Khaksari, K.; Soltanian-Zadeh, H. Automatic recognition of five types of white blood cells in peripheral blood. In Proceedings of the International conference image analysis and recognition, Póvoa de Varzim, Portugal, 21–23 June 2010; pp. 161–172.
10. Panagiotakis, C.; Argyros, A. Region-based Fitting of Overlapping Ellipses and its application to cells segmentation. *Image Vis. Comput.* **2020**, *93*, 103810. [\[CrossRef\]](#)

11. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
12. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
13. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
14. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [[CrossRef](#)]
15. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
16. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 3–11.
17. 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.
18. Wollmann, T.; Gunkel, M.; Chung, I.; Erfle, H.; Rippe, K.; Rohr, K. GRUU-Net: Integrated convolutional and gated recurrent neural network for cell segmentation. *Med. Image Anal.* **2019**, *56*, 68–79. [[CrossRef](#)] [[PubMed](#)]
19. Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; Wang, M. Swin-unet: Unet-like pure transformer for medical image segmentation. *arXiv* **2021**, arXiv:2105.05537.
20. Lux, F.; Matula, P. Cell segmentation by combining marker-controlled watershed and deep learning. *arXiv* **2020**, arXiv:2004.01607.
21. Fu, F.; Guenther, A.; Sakhdari, A.; McKee, T.D.; Xia, D. Deep learning accurately quantifies plasma cell percentages on CD138-stained bone marrow samples. *J. Pathol. Inform.* **2022**, *13*, 100011. [[CrossRef](#)] [[PubMed](#)]
22. Li, H.; Zhao, X.; Su, A.; Zhang, H.; Liu, J.; Gu, G. Color space transformation and multi-class weighted loss for adhesive white blood cell segmentation. *IEEE Access* **2020**, *8*, 24808–24818. [[CrossRef](#)]
23. Albawi, S.; Mohammed, T.A.; Al-Zawi, S. Understanding of a convolutional neural network. In Proceedings of the 2017 International Conference on Engineering and Technology (ICET, Antalya, Turkey, 21–23 August 2017; pp. 1–6.
24. Yang, Z.; Zhu, L.; Wu, Y.; Yang, Y. Gated channel transformation for visual recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 11794–11803.
25. Milletari, F.; Navab, N.; Ahmadi, S.A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 565–571.
26. Roy, A.G.; Navab, N.; Wachinger, C. Recalibrating fully convolutional networks with spatial and channel “squeeze and excitation” blocks. *IEEE Trans. Med. Imaging* **2018**, *38*, 540–549. [[CrossRef](#)] [[PubMed](#)]
27. Meyer, F. Topographic distance and watershed lines. *Signal Process.* **1994**, *38*, 113–125. [[CrossRef](#)]