

Article

# Breast Tumor Tissue Image Classification Using DIU-Net

Jiann-Shu Lee \* and Wen-Kai Wu

Department of Computer Science and Information Engineering, National University of Tainan,  
Tainan 700, Taiwan

\* Correspondence: cslee@mail.nutn.edu.tw

**Abstract:** Inspired by the observation that pathologists pay more attention to the nuclei regions when analyzing pathological images, this study utilized soft segmentation to imitate the visual focus mechanism and proposed a new segmentation–classification joint model to achieve superior classification performance for breast cancer pathology images. Aiming at the characteristics of different sizes of nuclei in pathological images, this study developed a new segmentation network with excellent cross-scale description ability called DIU-Net. To enhance the generalization ability of the segmentation network, that is, to avoid the segmentation network from learning low-level features, we proposed the Complementary Color Conversion Scheme in the training phase. In addition, due to the disparity between the area of the nucleus and the background in the pathology image, there is an inherent data imbalance phenomenon, dice loss and focal loss were used to overcome this problem. In order to further strengthen the classification performance of the model, this study adopted a joint training scheme, so that the output of the classification network can not only be used to optimize the classification network itself, but also optimize the segmentation network. In addition, this model can also provide the pathologist model’s attention area, increasing the model’s interpretability. The classification performance verification of the proposed method was carried out with the BreaKHis dataset. Our method obtains binary/multi-class classification accuracy 97.24/93.75 and 98.19/94.43 for 200× and 400× images, outperforming existing methods.



**Citation:** Lee, J.-S.; Wu, W.-K. Breast Tumor Tissue Image Classification Using DIU-Net. *Sensors* **2022**, *22*, 9838. <https://doi.org/10.3390/s22249838>

Academic Editors: Zongyuan Ge, Donghao Zhang, Deval Mehta and Dwarikanath Mahapatra

Received: 17 October 2022  
Accepted: 12 December 2022  
Published: 14 December 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/>).

**Keywords:** soft segmentation; DIU-Net; joint training

## 1. Introduction

Cancer is currently one of the leading causes of human death worldwide, and for women, breast cancer is the second main cause of cancer death after lung cancer [1]. According to the International Agency for Research on Cancer (IARC), which is part of the World Health Organization (WHO) [2], the number of deaths caused by cancer is expected to increase to more than 27 million by 2030 [3]. Commonly used breast cancer clinical screening methods include mammography [4], breast ultrasound [5], biopsy [6] and other radiology imaging. Radiology images can help to identify abnormal areas. However, they cannot be used to determine whether the area is cancerous. The biopsy [6], where tissue is taken and studied under a microscope to see if cancer is present, is the only diagnostic procedure that can confirm whether a suspicious area is cancerous. After completing the biopsy, the diagnosis will be proceeded by the pathologists, who examine the tissue under a microscope, looking for cancerous cells. Pathologists determine cancerous regions and malignancy degree [7,8] by visually examining the regularities of cell shapes and tissue distributions. Microscopic examination of histological slides by a pathologist for diagnosis is considered the gold standard for making a definite diagnosis [9]. However, traditional manual diagnosis requires a lot of effort for a pathologist. Whether it is due to insufficient diagnosis experience or inattention, it is prone to make wrong diagnoses using the manual way. In contrast, the automatic classification of pathological images using computer-aided diagnosis (CAD) [10] can not only improve diagnostic efficiency but also provide more objective and stable diagnostic results.

In [11], a database of breast cancer histopathology images, called BreakHis, was introduced by Spanhol et al. to overcome the problem of small data sets which is the main obstacle leading to the lack of development of a new analysis method. For texture features, popular textural descriptors were used, such as Local Binary Patterns (LBP) [12], Completed LBP (CLBP) [13], Local Phase Quantization (LPQ) [14], Grey-Level Co-occurrence Matrix (GLCM) [15], Threshold Adjacency Statistics [16], and Oriented fast and Rotated BRIEF (ORB) [17]. For classification, four different classifiers were used to assess the above feature sets including a 1-Nearest Neighbor (1-NN), Quadratic Discriminant Analysis (QDA), Support Vector Machines (SVM), and Random Forests. Chan et al. [18] calculated the fractal dimension of images of breast cancer slides and then classified them into benign or malignant images via SVM. Kahya et al. [19] presented an adaptive sparse support vector machine by combining the support vector machine with the weighted L1-norm to classify the breast cancer histopathology images. However, the performance of the above methods is limited due to the manual-based feature design methodology, which can be overcome by the convolutional neural network (CNN).

In the last decade, the CNN has achieved great success in image and video analysis and has received the attention of pathology image analysis researchers. Bardou et al. [20] compared two machine learning schemes for classifying breast cancer pathology images into benign and malignant. The first scheme is based on the extraction of a set of hand-crafted features encoded by two coding models, bag of words and locality constrained linear coding, and trained by support vector machines. The second scheme is based on the design of CNNs. The experimental results show that CNNs outperformed the handcrafted feature-based classifier. Motlagh et al. [21] presented Inception and ResNet architectures to discriminate microscopic cancerous images. They demonstrated an automatic framework for breast tumor detection and classification of its subtypes. The above two methods employing the existing CNN models for feature extraction and classification cannot adapt to the innate characteristics of pathological images; hence, the classification performance is naturally limited. Jiang et al. [22], considering the characteristics of histopathological images, designed a new CNN architecture for the classification of breast cancer histopathology images using the small SE-ResNet module, which is named the breast cancer histopathology image classification network (BHCNet). This model was used for the automatic classification of breast cancer histology images into benign and malignant and eight subtypes. The study by Filipczuk et al. [23] confirms that nucleus features can effectively classify benign and malignant breast cancer. Based on this concept, George et al. [24] proposed a nucleus feature extraction method utilizing a convolutional neural network for automated breast cancer detection. Non-overlapping nuclei patches are detected from the images first, and then CNN is employed to extract features. A feature fusion approach with a support vector machine classifier is used to classify breast tumor images. However, since the detection of nuclei patches is obtained by traditional image processing methods, the accuracy is not high enough, and the patch detection errors or locating errors will affect the subsequent classification performance.

Complex structures of the pathological image and significant variations in the morphology of the same type of nucleus within and across images make pathological image classification a challenging task. The aforementioned BHCNet [22] can achieve decent classification results; however, there still exists room for further improvement. A feasible way to enhance breast tumor classification performance is to drive the CNN model to focus more on the nucleus regions in which the cancerous characteristics are contained. Based on this consideration, we proposed a segmentation–classification joint training mechanism to enhance the BHCNet. The segmentation module is responsible for learning the Nucleus Focus Map (NFM) in which the nucleus region corresponds to higher weights. The Nucleus Focus Weighted Image (NFWI) is obtained by multiplying the NFM by the input image, and then the NFWI is input to BHCNet. In response to the inconsistent size of nuclei in tissue slice images, we proposed a Dilated Inception U-Net (DIU-Net) model with better cross-scale description ability, which enhances the performance of nuclei segmentation. To

adapt to the significant variations in the morphology of the same type of nucleus, this study proposed a Complementary Color Conversion Scheme (C3S) to enhance the generalization ability of the segmentation module. Since the area of nuclei in a pathological image is much smaller than the non-nucleus area, there exists a data imbalance phenomenon. This makes the trained segmentation module apt to pay more attention to the non-nucleus regions, resulting in segmentation bias. This study combined dice loss and focal loss to overcome this problem. The experimental results show that our method can indeed achieve better performance than the BHCNet [22]. In addition, our model has another advantage, that is, it can provide visual information of the model to the pathologist so that the pathologist can understand where the area concerned by the model is, and further enhance the pathologist's trust in this model.

In summary, the main contributions of our study are as follows:

- (1) A segmentation–classification joint training mechanism was proposed to enhance the classification performance for breast cancer pathology images.
- (2) A Dilated Inception U-Net model with better cross-scale description ability was proposed to enhance the performance of nuclei segmentation.
- (3) A Complementary Color Conversion Scheme was proposed to enhance the generalization ability of the nuclei segmentation.
- (4) The proposed model can provide visual information of the model to the pathologist so that the pathologist can understand where the area concerned by the model is.

The rest of the paper is organized as follows: Section 2 is dedicated to the proposed segmentation–classification joint model and its training. In Sections 3 and 4, the dataset, the implementation settings, performances and experimental results comparisons are given and discussed.

## 2. Method

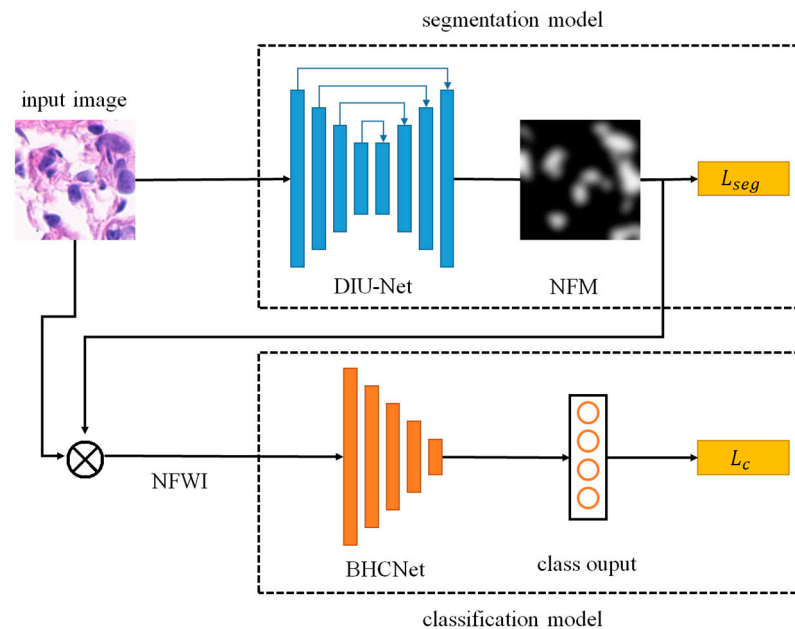
### 2.1. System Architecture

The architecture of the proposed method is shown in Figure 1, including two parts: a segmentation network and a classification network. The architecture can be trained using the segmentation–classification joint training mechanism to achieve good performance for breast cancer classification. After the input image is softly segmented by the segmentation network, NFM is output and  $L_{seg}$  is the corresponding segmentation loss. The NFWI is obtained by multiplying the NFM by the input image, and then the NFWI is classified by the classification network and the corresponding classification loss  $L_c$  is calculated. Detailed explanations of the different parts of the proposed network are provided in the following sections.

### 2.2. Segmentation Network

The nucleus features can be utilized to effectively classify benign and malignant breast cancer [23]. Pathologists pay more attention to the nucleus region when analyzing pathological images, so this study uses soft segmentation to imitate the visual focus mechanism of pathologists. Figure 2 shows the results of hard segmentation and soft segmentation for a pathological tissue patch exemplar in which the values of Figure 2c indicate the corresponding visual focus weights. To learn the corresponding NFM the segmentation network should possess pixel-level discrimination ability. U-net [25] was invented for semantic segmentation with an architecture that can be viewed as an encoder network followed by a decoder network. Unlike a classification network where the end result of the network is the only high-level semantic features, semantic segmentation not only requires discrimination at the pixel level but also a mechanism to project the discriminative features learned at different stages of the encoder onto the pixel space. U-Net is well recognized for its good performance in medical image segmentation tasks. To further enhance the segmentation performance of U-Net, some new architectures based on U-Net have been proposed in recent years. The most attractive one is the Attention U-Net [26]. Compared with R2U-Net [27] and U-Net++ [28], which need to use more parameters to improve

performance, Attention U-Net can significantly improve performance with relatively fewer parameters. Therefore, this study uses Attention U-Net as the basic architecture of the segmentation network.



**Figure 1.** System architecture of the proposed method. For the binary classification case, the class output corresponds to benign/malignant, while for the multi-classification case, the class output corresponds to the eight subcategories.



**Figure 2.** (a) A pathological tissue patch. (b) Hard segmentation result of (a). (c) Soft segmentation result of (a).

In fact, the size of nuclei in pathological images is not uniform, so the segmentation network must be able to capture cross-scale features to focus on nuclei of different sizes. Experiments have confirmed that the Inception structure has the ability to capture features at different scales [29], but unfortunately, the pooling layer in the Inception structure weakens this ability of the overall model due to the information loss coming from pooling. To overcome this shortcoming, this study replaces the pooling layer of Inception with dilated convolution [30], and calls the replaced block, as shown in Figure 3, Dilated Inception (DI). To further strengthen the cross-scale description capability of the Attention U-Net, we replace the convolutional layers of the encoder and decoder of the Attention U-Net with the DI blocks. The adjusted architecture is called DIU-Net as shown in Figure 4.

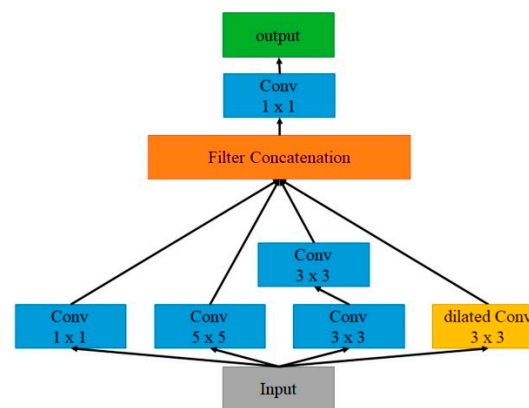


Figure 3. Dilated Inception block.

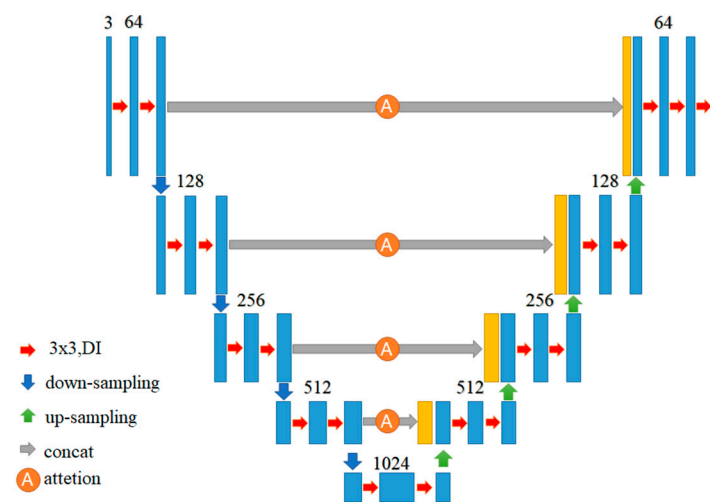


Figure 4. The architecture of the DIU-Net, where convolutional layers are replaced by DIs and the numbers indicate the number of channels.

To enhance the generalization ability of DIU-Net, it is necessary to avoid the operation of DIU-Net relying on the color information and relative brightness of the nucleus and the background which are less reliable. C3S was proposed to achieve this goal. By converting some training samples to complementary colors, DIU-Net learns to ignore the color and relative brightness of the nucleus and background and instead learns to use higher-level texture features for segmentation judgment. Pathological images are inherently characterized by the fact that the area of nuclei is much smaller than the background area, which leads to data imbalance. If an inappropriate loss function is used, it is easy for the trained segmentation network to pay more attention to the accuracy of non-nucleus regions, resulting in segmentation bias. Dice loss belongs to region-based loss and uses the relative overlap rate between the prediction result and the ground truth to quantify the loss. It is not affected by the size of the target object innately, so it is suitable for the loss of the segmentation network. However, using dice loss alone is prone to instability when the relative overlap rate between the prediction result and the ground truth is zero during the training process. To avoid this problem, this study added focal loss which belongs to distribution-based loss and can also help improve the problem of data imbalance. The dice loss is defined in Equation (1), where  $N$  is the total pixel number in one batch,  $\hat{y}_i$  and  $y_i$  denote the prediction result and the ground truth for pixel  $i$ , respectively. The  $\epsilon$  term is

used to ensure the loss function stability by avoiding the numerical issue of dividing by 0. The focal loss is defined in Equation (3).

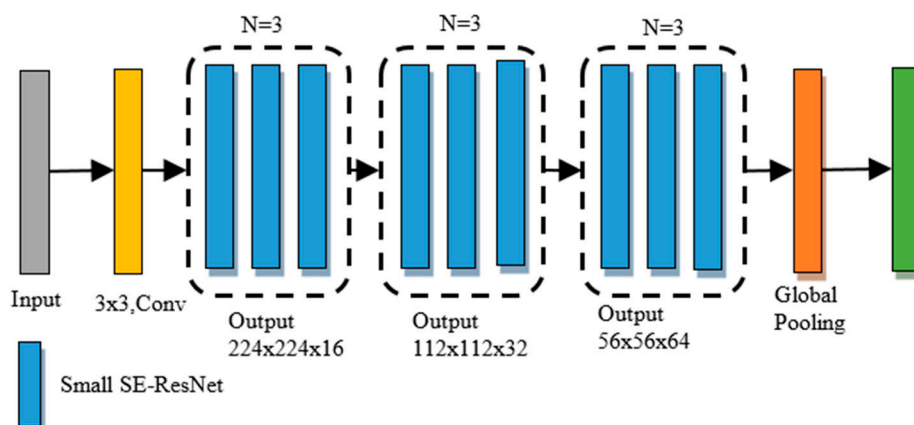
$$L_{dice} = 1 - \frac{\sum_{i=1}^N y_i \hat{y}_i + \epsilon}{\sum_{i=1}^N y_i + \hat{y}_i + \epsilon} - \frac{\sum_{i=1}^N (1 - y_i)(1 - \hat{y}_i) + \epsilon}{\sum_{i=1}^N 2 - y_i - \hat{y}_i + \epsilon} \quad (1)$$

$$\hat{y}_i = \begin{cases} \hat{y}_i & \text{if } y_i = 1 \\ 1 - \hat{y}_i & \text{otherwise} \end{cases} \quad (2)$$

$$L_{focal} = - \sum_{i=1}^N (1 - \hat{y}_i)^2 \log(\hat{y}_i) \quad (3)$$

### 2.3. Classification Network

In this study, BHCNet [22], which currently has outstanding breast tumor classification performance in the public dataset BreakHis, was used as the classification network, and its architecture is shown in Figure 5. BHCNet-3 was used for benign and malignant tumor classification tasks. For the more difficult subtypes classification task, BHCNet-6 was used. The difference between the two networks is the network depth. Since the subtypes classification task is more difficult than the binary classification task, a deeper network is required to cope with it.



**Figure 5.** The architecture of the BHCNet-3.

### 2.4. Training

The loss function of the overall network can be divided into two parts: segmentation loss  $L_{seg}$  and classification loss  $L_c$ . The definition of segmentation loss  $L_{seg}$  is defined in Equation (4) where  $\lambda_s$  is the weight of focal loss.  $L_{seg}$  shoulders our expectation for the output of the segmentation network, that is, to appropriately reflect the degree of attention to the nucleus in the image. The loss of benign/malignant binary classification is denoted as  $L_{bce}$ , which is defined in Equation (5), where  $M$  is the total image number in one batch for the binary classification task,  $\hat{y}_k$  and  $y_k$  denote the prediction result and the ground truth for image  $k$ , respectively. As for the loss of subcategory classification, it is denoted as  $L_{mce}$ , which is defined in Equation (6), where  $K$  is the total image number in one batch for a multi-class classification task,  $\hat{y}_m$  represents the target class prediction result of image  $m$ . The total loss  $L_{total}$  is defined in Equation (8) where  $\lambda_c$  is the weight of classification loss  $L_c$ .

The training samples of the overall network in the training phase are partly from the segmentation training set and partly from the classification training set. The segmentation network and the classification network are jointly trained by the training samples of the two training sets. The training sample  $I_s$  from the segmentation training set is fed into DIU-Net to obtain NFM. The corresponding  $L_{seg}$  is calculated through the ground truth of  $I_s$ , and then the weights of DIU-Net are corrected via  $L_{seg}$ . The training sample  $I_c$  from the classification training set is fed into DIU-Net to obtain NFM. After multiplying NFM and  $I_c$  to obtain NFWI, then input to BHCNet to obtain classification prediction. The

corresponding loss  $L_c$  is calculated through the ground truth of  $I_c$ . Correcting the weights of BHCNet and DIU-Net via  $L_{total}$ :

$$L_{seg} = L_{dice} + \lambda_s L_{focal} \quad (4)$$

$$L_{bce} = - \sum_{k=1}^M [y_k \log(\hat{y}_k) + (1 - y_k) \log(1 - \hat{y}_k)] \quad (5)$$

$$L_{mce} = - \sum_{m=1}^K \log(\hat{y}_m) \quad (6)$$

$$L_c = \begin{cases} L_{bce}, & \text{two classes} \\ L_{mce}, & \text{multi-class} \end{cases} \quad (7)$$

$$L_{total} = (1 - \lambda_c) * L_{seg} + \lambda_c * L_c \quad (8)$$

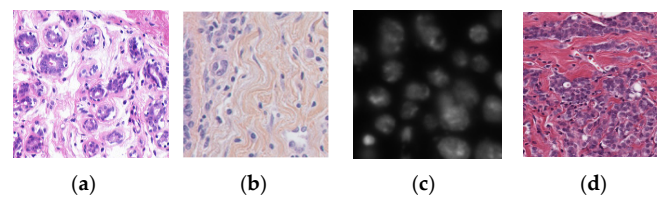
### 3. Experiments

In this study, accuracy (ACC) and dice coefficient were used as efficacy evaluation indicators. The segmentation experiment is divided into two parts: the training process with/without the C3S, in order to understand the benefits of this scheme. As for the verification method of the classification experiment, the same evaluation methods were adopted as other studies [20,22,24,31,32] using the BreakHis dataset.

#### 3.1. Datasets

Since few of the currently public breast cancer image datasets provide both the ground truth for nucleus segmentation and the ground truth for tumor types, the dataset used in this study is divided into a segmentation dataset and a classification dataset. The images of the segmentation dataset were collected from 4 sub-datasets, namely the UCSB dataset [33], the TNBC dataset [34], the 2018DSB dataset [35] and the MoNuSeg dataset [36]. There are 58 breast cancer images in the UCSB dataset, the image format is TIF, the resolution is  $896 \times 768$ , and the staining colors are relatively consistent. The TNBC dataset has 50 breast cancer images in PNG format with a resolution of  $512 \times 512$ , and their staining colors are relatively inconsistent. The 2018DSB dataset has 670 images of Spot nuclei, the image format is TIF, the resolution is  $256 \times 256$  to  $1388 \times 1040$  and the staining colors are also relatively inconsistent. The MoNuSeg dataset has 30 images. The tissue images are from patients with breast cancer, liver cancer, kidney cancer, prostate cancer, bladder cancer, colon cancer and gastric cancer. Since this dataset covers cells of different tissues, and the cell characteristics are highly variable, it is quite suitable for verifying the generalization ability of the segmentation network.

In terms of the appearance of staining, the images of the above four sub-datasets can be divided into two types: cells with a darker color than the background (referred to as the darker type) and cells with a lighter color than the background (referred to as the lighter type). Most of the images in the three sub-datasets UCSB, MoNuSeg and TNBC are of the darker type, while the 2018DSB sub-dataset is of the lighter type. Figure 6 shows sample images of each sub-dataset. In order to test the generalization ability of the segmentation network, this study made special arrangements for the training and testing datasets. UCSB and TNBC were used as segmentation training datasets, while MoNuSeg and 2018DSB were used as segmentation test datasets. A total of 80% of the data in the segmentation training dataset is used as training data, and the remaining 20% is used as validation data. Since the characteristics of the test data are quite different from the training data, if the test performance is good, it can be confirmed that the segmentation network model proposed in this study has excellent generalization ability.



**Figure 6.** Sample images of four sub-datasets for segmentation: (a) UCSB; (b) TNBC; (c) 2018DSB; (d) MoNuSeg.

The classification dataset is the BreakHis dataset, which uses H&E staining, the image size is  $700 \times 400$ , and the image magnifications are  $200\times$  and  $400\times$ . Each image has a benign/malignant label and the corresponding subcategory label, for a total of 3833 images. The subcategories of benign tumors are adenosis (A), fibroadenoma (F), phyllodes tumor (PT) and tubular adenoma (TA). The malignant tumor subcategories are ductal carcinoma (DC), lobular carcinoma (LC), mucinous carcinoma (MC) and papillary carcinoma (PC). There are 2013  $200\times$  images, which are called the  $200\times$  dataset, and 1820  $400\times$  images, which are called the  $400\times$  dataset. Randomly select 70% from the  $200\times$  dataset as the  $200\times$  training set and the remaining 30% as the  $200\times$  test set. Likewise, randomly select 70% from the  $400\times$  dataset as the  $400\times$  training set and the remaining 30% as the  $400\times$  test set. The union of the  $200\times$  training set and the  $400\times$  training set is called the classification training set, and the union of the  $200\times$  test set and the  $400\times$  test set is called the classification test set.

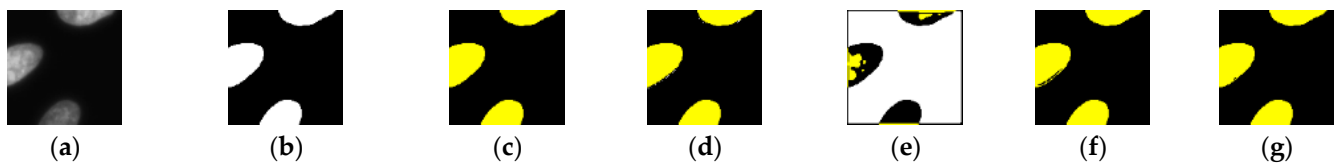
### 3.2. Evaluating the Segmentation Performance

U-Net [25] can be regarded as a representative model for medical image segmentation. In recent years, some improved architectures based on U-Net have been proposed, such as Attention U-Net (Att-U-Net) [26], R2U-Net [27] and U-Net++ [28]. These models all have quite good segmentation performance. This experiment compares the segmentation performance of the proposed DIU-Net with these models under the same training and testing conditions. Table 1 shows the test results of the training process of each model without using the C3S for training. The results show that our proposed model outperforms the other methods on both the validation set and the test set and R2U-Net has the worst segmentation effect. However, the performance of all models on the test set is not ideal. The main reason is that these models learn to rely on the color and relative brightness of the nucleus and background to make judgments during the training process, which reduces the performance of the test data with different characteristics. Figure 7 shows examples of the segmentation results of the 2018 DSB sub-dataset. It can be seen that all segmentation models cannot correctly segment the nucleus region. Table 2 shows the test results of the training process of each model using the C3S. It can be found that the segmentation performance of all models is greatly improved, indicating that the C3S can indeed improve the generalization ability of these models. The experimental results also show that our proposed model performs better than other methods in both the validation set and test set, and the segmentation effect of R2U-Net is still the worst. The actual segmentation performance can refer to the segmentation examples in Figure 8. Compared with Figure 7, the segmentation performance is indeed significantly improved.

**Table 1.** The test results of different segmentation models without C3S in the training phase.

	U-Net	Att-U-Net	R2U-Net	U-Net++	Ours
Validation Set					
Dice Coefficient	0.69	0.68	0.35	0.68	<b>0.70</b>
Test Set					
Dice Coefficient	0.39	0.40	0.36	0.36	<b>0.41</b>

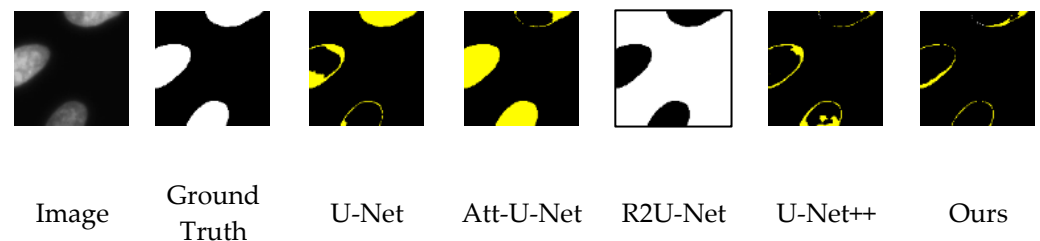




**Figure 7.** The segmentation results using different models without the C3S for a sample image of the 2018 DSB sub-dataset. The white pixel represents FP (false positive), while the yellow pixel represents FN (false negative). (a) Image; (b) Ground Truth; (c) U-Net; (d) Att-U-Net; (e) R2U-Net; (f) U-Net++; (g) Ours.

**Table 2.** The test results of different segmentation models with C3S in the training phase.

	U-Net	Att-U-Net	R2U-Net	U-Net++	Ours
Validation Set					
Dice Coefficient	0.84	0.84	0.24	0.85	<b>0.85</b>
Test Set					
Dice Coefficient	0.50	0.53	0.33	0.52	<b>0.54</b>



**Figure 8.** The segmentation results using different models with the C3S during the training phase for the same image in Figure 7.

To further understand the impact of different loss and C3S on the performance of the segmentation network, this part uses the performance of the validation set and the test set to conduct ablation experiments, and the results are shown in Table 3. It can be seen from the results that the performance of using only dice loss is worse than that of using focal loss, while the performance of combining dice loss with focal loss is the best. Furthermore, no matter what kind of loss is used, as long as the C3S is used in the training process, better segmentation performance can be obtained. The results show that the combination of dice loss and focal loss is beneficial to the optimization of the segmentation network. The C3S used in the training process can improve the generalization ability of the segmentation network.

**Table 3.** Ablation experiments on the segmentation network.

Focal Loss	Dice Loss	Dice Coefficient (without C3S)	Dice Coefficient (with C3S)
Validation Set			
✓	×	0.66	0.84
×	✓	0.66	0.82
✓	✓	<b>0.70</b>	<b>0.85</b>
Test Set			
✓	×	0.41	0.47
×	✓	0.34	0.44
✓	✓	<b>0.41</b>	<b>0.54</b>

### 3.3. Evaluating the Classification Performance

Table 4 shows the results of the dichotomous classification of benign and malignant breast pathological images by various methods. CNN [20] means the binary/multi-class classification results using the CNN-based method of [20]. ResNet [21] means the binary/multi-class classification results using Inception and ResNet architectures proposed by [21]. BHCNet [22] is the baseline for classification performance comparison since our proposed method employs it as a classification module. NucDeep [24] means the binary/multi-class classification results using the method proposed by [24], in which a feature fusion approach with a support vector machine is used to classify breast tumor images. ResHist [31] is a residual learning-based CNN with 152 layers developed for breast tumor binary classification. myResNet-34 [32] is also a residual learning-based CNN derived from ResNet-34 via merging shallow features and using Leaky ReLU and Batch Normalization to enhance the malignancy-and-benign classification performance. The results show that the model proposed in this study has a higher classification accuracy than other methods whether in the 200× test set, 400× test set or the overall test set. The performance of various methods in the classification of benign and malignant subcategories of breast pathology images is shown in Table 5. The results show that the proposed model still has higher subcategory classification accuracy than other methods whether it is in the 200X test set, 400X test set or in the overall test set. This shows that the proposed strategy combining the degree of nucleus attention can indeed effectively improve classification performance. The multi-class confusion matrix of the proposed model is shown in Figure 9. From the confusion matrix, it can be found that LC is the most difficult category to be classified.

**Table 4.** Binary classification accuracy comparison.

Methods	200×	400×	Total
CNN [20]	96.36	95.97	96.16
ResNet [21]	93.64	93.16	93.4
BHCNet [22]	97.2	96.96	97.04
NucDeep [24]	96.21	×	96.21
ResHist [31]	91.15	86.27	×
myResNet-34 [32]	90.47	88.79	×
<b>Ours</b>	<b>97.24</b>	<b>98.19</b>	<b>97.74</b>

**Table 5.** Multi-class classification accuracy comparison.

Methods	200×	400×	Total
CNN [20]	80.83	81.03	80.93
ResNet [21]	76.54	79.58	78.06
BHCNet [22]	92.27	91.15	91.71
NucDeep [24]	63.3	×	63.3
<b>Ours</b>	<b>93.75</b>	<b>94.43</b>	<b>94.09</b>

	PC	MC	LC	DC	TA	PT	F	A
PC	40	0	0	0	0	1	0	0
MC	0	54	3	1	0	0	0	0
LC	0	3	35	11	0	0	0	0
DC	2	1	4	261	0	0	0	0
TA	0	0	0	0	39	0	2	1
PT	0	0	0	0	0	30	3	0
F	0	0	0	0	1	5	74	0
A	0	0	0	0	0	0	0	34

(a)

	PC	MC	LC	DC	TA	PT	F	A
PC	37	1	0	3	0	0	1	0
MC	0	48	0	1	1	0	0	1
LC	0	1	33	7	0	0	0	1
DC	0	1	4	231	1	0	0	0
TA	0	0	0	0	34	0	0	0
PT	0	0	0	0	0	31	2	2
F	0	0	0	0	0	1	69	2
A	0	0	0	0	0	0	0	32

(b)

**Figure 9.** The multi-class confusion matrix of the proposed model for (a) 200× test set and (b) 400× test set.

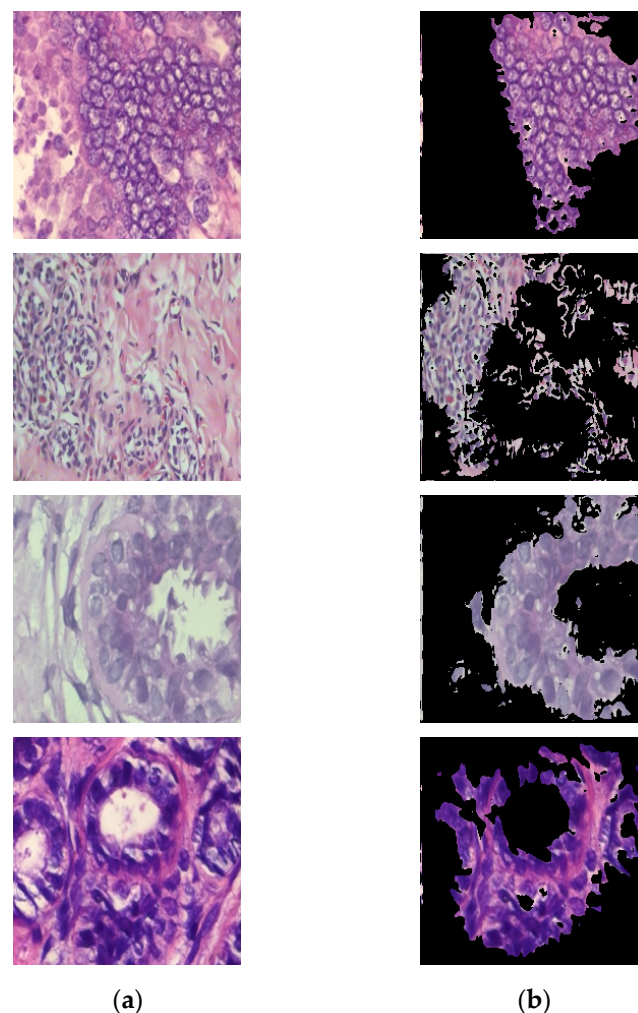
### 3.4. Ablation Study

In this ablation experiment, we want to understand the impact of several key factors on classification accuracy. The ablation experimental results are shown in Table 6. Omitting the C3S, the performance degradation is 0.56 and 1.42 for binary classification and multi-class classification, respectively. The results reveal that the improvement of segmentation performance by the C3S also indirectly boosts the classification performance. To understand the influence level of replacing the pooling layer of Inception with dilated convolution, the results with and without DI were compared. From Table 6, we can find that using DI the classification accuracy can be enhanced by 0.55 and 2.38 for binary and multi-class classification, respectively. This reflects that DI can indeed improve the cross-scale description ability of Attention U-Net for the nucleus in breast pathology images. To further understand the impact of the joint training of the segmentation network and the classification network on the accuracy of the binary classification and multi-class classification, a control version, the separated training version, was designed to compare the classification performance of the two. This version trains the segmentation network and the classification network separately and then concatenates them after the training is complete. The comparison results are shown in Table 6. The classification performance of the joint training version outperforms the separated training versions by 6.32 and 18.11 on the binary and multi-class tasks, respectively. This substantial performance improvement is due to two factors. One is that the soft segmentation results of the segmentation network allow the classification network to focus more on the nucleus regions. The other is that the joint training scheme gives classification results, which can not only be fed back to optimize the classification network, but also optimize the segmentation network to generate the corresponding NFM that is more conducive to classification performance.

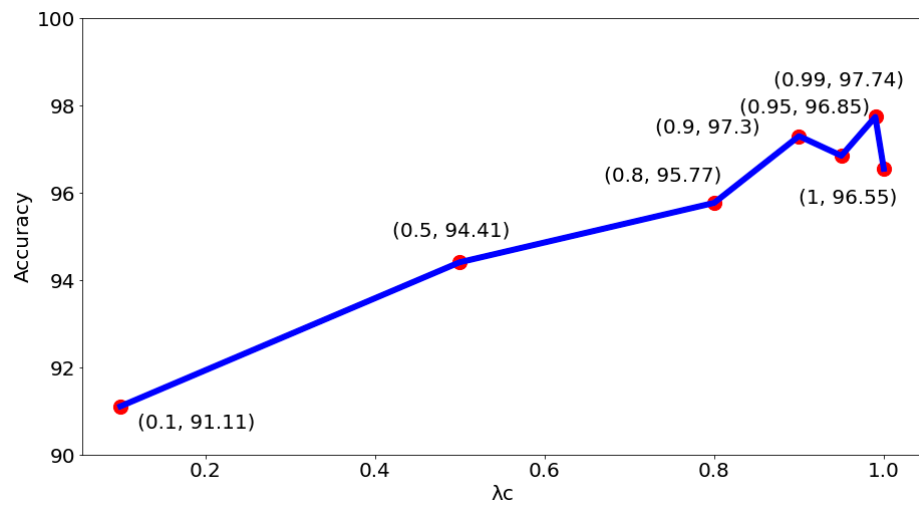
**Table 6.** Ablation experiment on classification accuracy.

Methods	Binary Classification	Multi-Class Classification
Without C3S	97.18	92.67
Without DI	97.04	91.71
Separated training version	91.42	75.98
Ours	97.74	94.09

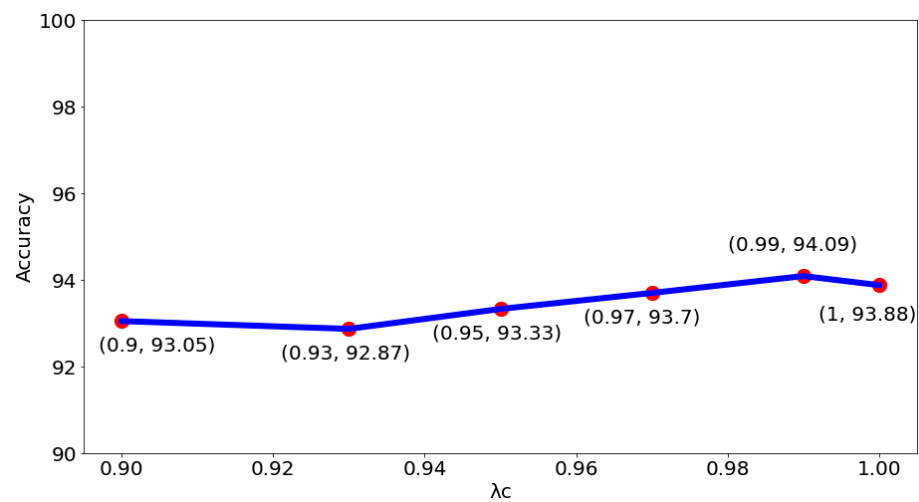
In addition to yielding superior classification results, the proposed method has the side benefit of being able to provide model visualization. That is, the areas concerned by the model can be displayed to the pathologist for reference (as shown in Figure 10), thereby enhancing the physician's confidence in the classification results. It can also be found from Figure 10 that the segmentation results of the proposed method cover almost all the nucleus regions. This also shows the effectiveness of the proposed method for nucleus segmentation. To understand the impact of choosing different  $\lambda_c$  on the system performance, tests for different  $\lambda_c$  were proceeded. Figure 11 shows the test results of the binary classification task. The results show that when  $\lambda_c$  is set to 0.99, there is the best classification performance. As for the test results of the subcategory classification task, as shown in Figure 12, the results also show that the best classification performance is obtained when  $\lambda_c$  is set to 0.99. This means that the hyperparameter  $\lambda_c$  of the proposed model is quite stable, whether facing binary or subcategory tasks. This stable property indirectly reflects the value of the proposed model in clinical applications.



**Figure 10.** The segmentation results of the proposed method: (a) the input images; (b) segmented results of our method.



**Figure 11.** The impact of choosing different  $\lambda_c$  on accuracy for binary classification.



**Figure 12.** The impact of choosing different  $\lambda_c$  on accuracy for multi-class classification.

#### 4. Conclusions

Inspired by pathologists' interpretation of pathological images, this study proposed an automatic classification model of breast pathological images that combines a segmentation network and classification network. The soft segmentation results are generated by the segmentation network to simulate the pathologist's relative attention to the viewing mode of the nucleus regions, and then the weighted images are input into the classification network for classification. Considering the phenomenon of different sizes of nuclei in pathological images, the DIU-Net proposed in this study has excellent cross-scale description ability, so that nuclei of different sizes can correspond to higher attention coefficients. In addition, to make the segmentation network have better generalization ability, C3S was used in the training phase to guide the segmentation network to avoid learning low-level features such as color or relative brightness and instead learn higher-level texture features to identify nuclei. With the background, the generalization ability of the segmentation network has been successfully improved. Furthermore, dice loss and focal loss were used to successfully overcome the data imbalance caused by the disparate area ratio of nuclei to the background. This study utilized a joint training scheme so that the output of the classification network can not only be used to optimize the classification network itself, but also optimize the segmentation network to further strengthen the classification performance of the model. The experimental results show that the model proposed in this study outperforms the existing classification models for both  $200\times$  and  $400\times$  pathological images in both binary

and subcategory classification tasks. In addition, the proposed model has the side benefit of being able to provide the areas concerned by the model. This visualization can enhance the pathologist's confidence in the model's classification results. The greatest contribution of the model developed in this study is that it can provide pathologists with excellent classification results and at the same time provide information on the area of interest of the model, thereby assisting pathologists in making decisions. The main drawback of this model is that it cannot be applied to small-magnification pathological images because this model must first softly segment the nucleus area, and small-magnification pathological images cannot clearly reveal the nucleus.

**Author Contributions:** Conceptualization, J.-S.L.; Data curation, W.-K.W.; Funding acquisition, J.-S.L.; Methodology, J.-S.L. and W.-K.W.; Software, W.-K.W.; Supervision, J.-S.L.; Writing—original draft, W.-K.W.; Writing—review and editing, J.-S.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Ministry of Science and Technology under Grant MOST 108-2221-E-024-011-MY3.

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

## References

1. Chun, M.C. Breast Cancer: Symptoms, Risk Factors, and Treatment, Medical News Today. 2018. Available online: <https://www.medicalnewstoday.com/articles/37136.php> (accessed on 10 March 2018).
2. World Health Organization. Available online: <http://www.who.int/en/> (accessed on 10 March 2018).
3. Boyle, P.; Levin, B. World Cancer Report. 2008. Available online: <https://publications.iarc.fr/Non-Series-Publications/World-Cancer-Reports/World-Cancer-Report-2008> (accessed on 16 October 2022).
4. Lim, C.N.; Suliong, C.; Rao, C.V.; Aung, T.; Sieman, J.; Tin, W. Recent Advances in Breast Cancer Diagnosis Entering an Era of Precision Medicine. *Borneo J. Med. Sci. (BJMS)* **2019**, *13*. [[CrossRef](#)]
5. Chen, D.R.; Lin, Y.C.; Huang, Y.L. 3D Contouring for Breast Tumor in Sonography. *arXiv* **2019**, arXiv:190109407.
6. Horvat, J.V.; Keating, D.M.; Rodrigues-Duarte, H.; Morris, E.A.; Mango, V.L. Calcifications at Digital Breast Tomosynthesis: Imaging Features and Biopsy Techniques. *RadioGraphics* **2019**, *39*, 307. [[CrossRef](#)] [[PubMed](#)]
7. He, L.; Long, L.R.; Antani, S.; Thoma, G. Computer assisted diagnosis in histopathology, Sequence Genome Anal. *Methods Appl.* **2010**, *3*, 271–287.
8. He, L.; Long, L.R.; Antani, S.; Thoma, G.R. Histology image analysis for carcinoma detection and grading. *Comput. Methods Programs Biomed.* **2012**, *107*, 538–556. [[CrossRef](#)]
9. Das, K.; Conjeti, S.; Roy, A.G.; Chatterjee, J.; Sheet, D. Multiple instance learning of deep convolutional neural networks for breast histopathology whole slide classification. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI), Washington, DC, USA, 4–7 April 2018; pp. 578–581.
10. Araújo, T.; Aresta, G.; Castro, E.; Rouco, J.; Aguiar, P.; Eloy, C.; Campilho, A. Classification of breast cancer histology images using convolutional neural networks. *PLoS ONE* **2017**, *12*, e0177544. [[CrossRef](#)]
11. Spanhol, F.A.; Oliveira, L.S.; Petitjean, C.; Heutte, L. A dataset for breast cancer histopathological image classification. *IEEE Trans. Biomed. Eng.* **2016**, *63*, 1455–1462. [[CrossRef](#)]
12. Ojala, T.; Pietikainen, M.; Maenpää, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*; IEEE Computer Society: Washington, DC, USA, 2002; Volume 24, pp. 971–987.
13. Guo, Z.; Zhang, L.; Zhang, D. A completed modeling of local binary pattern operator for texture classification. *IEEE Trans. Image Process.* **2010**, *19*, 1657–1663.
14. Ojansivu, V.; Heikkilä, J. Blur insensitive texture classification using local phase quantization. In Proceedings of the 3rd International Conference on Image and Signal Processing (ICISP), Cherbourg-Octeville, France, 1–3 July 2008; Volume 5099, pp. 236–243.
15. Haralick, R.; Shanmugam, K.; Dinstein, I. Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *3*, 610–621. [[CrossRef](#)]
16. Hamilton, N.A.; Pantelic, R.S.; Hanson, K.; Teasdale, R.D. Fast automated cell phenotype image classification. *BMC Bioinform.* **2007**, *8*, 110. [[CrossRef](#)]
17. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An efficient alternative to SIFT or SURF. In Proceedings of the IEEE International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 2564–2571.
18. Chan, A.; Tuszynski, J.A. Automatic prediction of tumour malignancy in breast cancer with fractal dimension. *Open Sci.* **2016**, *3*, 160558. [[CrossRef](#)] [[PubMed](#)]

19. Kahya, M.A.; Al-Hayani, W.; Algamal, Z.Y. Classification of breast cancer histopathology images based on adaptive sparse support vector machine. *J. Appl. Math. Bioinform.* **2017**, *7*, 49.
20. Bardou, D.; Zhang, K.; Ahmad, S.M. Classification of Breast Cancer Based on Histology Images Using Convolutional Neural Networks. *IEEE Access* **2018**, *6*, 24680–24693. [[CrossRef](#)]
21. Jannesari, M.; Habibzadeh, M.; Aboulkheyr, H.; Khosravi, P.; Elemento, O.; Totonchi, M.; Hajirasouliha, I. Breast Cancer Histopathological Image Classification: A Deep Learning Approach. In Proceedings of the 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, 3–6 December 2018; pp. 2405–2412.
22. Jiang, Y.; Chen, L.; Zhang, H.; Xiao, X. Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module. *PLoS ONE* **2019**, *14*, e0214587. [[CrossRef](#)] [[PubMed](#)]
23. Adeshina, S.A.; Adedigba, A.P.; Adeniyi, A.A.; Aibinu, A.M. Breast cancer histopathology image classification with deep convolutional neural networks. In Proceedings of the 2018 14th International Conference on Electronics Computer and Computation (ICECCO), Kaskelen, Kazakhstan, 29 November–1 December 2018; pp. 206–212.
24. George, K.; Sankaran, P.; Joseph, K. Computer assisted recognition of breast cancer in biopsy images via fusion of nucleus-guided deep convolutional features. *Comput. Methods Programs Biomed.* **2020**, *194*, 105531. [[CrossRef](#)]
25. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional Networks for Biomedical Image Segmentation. In *2015 International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2015; pp. 234–241.
26. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Rueckert, D. Attention U-Net: Learning where to look for the pancreas. *arXiv* **2018**, arXiv:1804.03999.
27. Alom, M.Z.; Hasan, M.; Yakopcic, C.; Taha, T.M.; Asari, V.K. Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation. *arXiv* **2018**, arXiv:1802.06955.
28. 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.
29. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
30. Yu, F.; Koltun, V. Multi-scale context aggregation by dilated convolutions. In Proceedings of the 4th International Conference on Learning Representations, San Juan, Puerto Rico, 2–4 May 2016.
31. Gour, M.; Jain, S.; Kumar, T.S. Residual learning based CNN for breast cancer histopathological image classification. *Int. J. Imaging Syst. Technol.* **2020**, *30*, 621–635. [[CrossRef](#)]
32. Hu, C.; Sun, X.; Yuan, Z.; Wu, Y. Classification of breast cancer histopathological image with deep residual learning. *Int. J. Imaging Syst. Technol.* **2021**, *31*, 1583–1594. [[CrossRef](#)]
33. Gelasca, E.D.; Byun, J.; Obara, B.; Manjunath, B.S. Evaluation and benchmark for biological image segmentation. In Proceedings of the 2008 15th IEEE International Conference on Image Processing, San Diego, CA, USA, 12–15 October 2008.
34. Naylor, P.; Lae, M.; Reyat, F.; Walter, T. Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE Trans. Med. Imaging* **2018**, *38*, 448–459. [[CrossRef](#)]
35. Booz Allen Hamilton. 2018 Data Science Bowl. Kaggle. Available online: <https://www.kaggle.com/c/data-science-bowl-2018> (accessed on 16 October 2022).
36. Kumar, N.; Verma, R.; Sharma, S.; Bhargava, S.; Vahadane, A.; Sethi, A. A Dataset and a Technique for Generalized Nuclear Segmentation for Computational Pathology. *IEEE Trans. Med. Imaging* **2017**, *36*, 1550–1560. [[CrossRef](#)] [[PubMed](#)]