

Article

# MurSS: A Multi-Resolution Selective Segmentation Model for Breast Cancer

Joonho Lee <sup>1</sup>, Geongyu Lee <sup>1</sup>, Tae-Yeong Kwak <sup>1</sup>, Sun Woo Kim <sup>1</sup>, Min-Sun Jin <sup>2,†</sup>, Chungyeul Kim <sup>3,†</sup>  
and Hyeyoon Chang <sup>1,\*,†</sup>

<sup>1</sup> Deep Bio Inc., Seoul 08380, Republic of Korea; joonho.lee@deepbio.co.kr (J.L.); gglee@deepbio.co.kr (G.L.); tykwak@deepbio.co.kr (T.-Y.K.); swkim@deepbio.co.kr (S.W.K.)

<sup>2</sup> Department of Pathology, Bucheon St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul 14647, Republic of Korea; seasy@hanmail.net

<sup>3</sup> Department of Pathology, Korea University Guro Hospital, Seoul 08308, Republic of Korea; idea1@hanmail.net

\* Correspondence: hychang@deepbio.co.kr

† These authors contributed equally to this work.

**Abstract:** Accurately segmenting cancer lesions is essential for effective personalized treatment and enhanced patient outcomes. We propose a multi-resolution selective segmentation (MurSS) model to accurately segment breast cancer lesions from hematoxylin and eosin (H&E) stained whole-slide images (WSIs). We used The Cancer Genome Atlas breast invasive carcinoma (BRCA) public dataset for training and validation. We used the Korea University Medical Center, Guro Hospital, BRCA dataset for the final test evaluation. MurSS utilizes both low- and high-resolution patches to leverage multi-resolution features using adaptive instance normalization. This enhances segmentation performance while employing a selective segmentation method to automatically reject ambiguous tissue regions, ensuring stable training. MurSS rejects 5% of WSI regions and achieves a pixel-level accuracy of 96.88% (95% confidence interval (CI): 95.97–97.62%) and mean Intersection over Union of 0.7283 (95% CI: 0.6865–0.7640). In our study, MurSS exhibits superior performance over other deep learning models, showcasing its ability to reject ambiguous areas identified by expert annotations while using multi-resolution inputs.



**Citation:** Lee, J.; Lee, G.; Kwak, T.-Y.; Kim, S.W.; Jin, M.-S.; Kim, C.; Chang, H. MurSS: A Multi-Resolution Selective Segmentation Model for Breast Cancer. *Bioengineering* **2024**, *11*, 463. <https://doi.org/10.3390/bioengineering11050463>

Academic Editor: Zeynettin Akkus

Received: 29 March 2024

Revised: 2 May 2024

Accepted: 3 May 2024

Published: 7 May 2024



**Copyright:** © 2024 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:** multi-resolution; segmentation; selective segmentation method; breast cancer

## 1. Introduction

Breast cancer is a prevalent cancer among women worldwide [1]. In the United States, breast cancer is the second leading cause of cancer-related deaths in women [2]. Breast cancer is diagnosed by analyzing pathological features such as tumor growth patterns and cytologic characteristics. The most common type is invasive ductal carcinoma (IDC). IDC accounts for approximately 70–80% of all breast cancers [3]. Accurately measuring invasive tumor size is essential in breast cancer prognosis. However, in cancer lesions, IDC may coexist with ductal carcinoma in situ (DCIS), sometimes showing intermingled patterns. Therefore, measuring invasive tumor size via microscopic examination may be challenging. The presence of an extensive intraductal component is a critical factor in predicting breast cancer prognosis. The rapid growth of digital pathology has opened up possibilities for AI to aid pathologists in diagnosis. Tumor area estimation techniques can assist pathologists by providing more accurate measurements of invasive tumors [4–6].

Due to their gigapixel size and high resolution, whole-slide images (WSIs) require high computational costs. For example, WSIs from The Cancer Genome Atlas (TCGA) dataset [7], which are widely used in pathological image analysis, have a resolution that is a gigapixel in size, significantly higher than that of standard image datasets. As a result, each WSI is divided into smaller pieces, which are then used to train deep learning models. In

this case, we need to consider the receptive field, which is the area of the input image that influences the output [8]. When extracting a patch at a higher resolution, the contextual information from the wider area is not utilized as the tissue around the patch is not included in the receptive field. This fact limits the use of high-resolution content information and omits low-resolution contextual information.

In computer vision, previous deep learning models have primarily been designed for high-resolution, singular images, as these images are typically compact and do not require tiling into smaller patches. The authors of [9] proposed U-Net to accurately segment cellular structures from medical image backgrounds. U-Net uses low-level features during the high-to-low downsampling process and combines them into same-resolution high-level features during the low-to-high upsampling process through skip connections. Additionally, U-Net employs single-resolution images as its input. The authors of [10] proposed DeepLabV3, which utilizes dilated convolution to expand the receptive field, thereby leveraging the broader contextual information. Most segmentation tasks compress high-resolution images by passing them through CNN layers and then employing an upsampling method. However, the authors of [11] proposed HRNet, a high-resolution network that maintains high resolutions throughout the entire process and progressively adds feature maps of lower resolutions in parallel as the layers deepen; this technique is described as multi-resolution streams in parallel. As a result, HRNet can accurately segment regions, even in complex images, since both high- and low-resolution features are combined. Various models are being developed that utilize connections between features of different scales and expand the receptive field, aiming to work with multi-scale feature maps or high-resolution images. Nowadays, research focuses on combining feature information from images of various resolutions. The authors of [12] argued that more detailed areas can be observed as the resolution increases, but overall morphological features, such as the background, are missed. The authors of [13] proposed an image cascade network (ICNet) using cascade feature fusion. ICNet is trained by designing networks for three different-resolution images and combining them. In this way, the critical feature information from each resolution is used. The authors of [14] proposed a deep multi-magnification network (DMMN) for multi-class breast cancer image segmentation consisting of three U-Nets. They placed three images of different resolutions in the encoder of U-Net, concatenating the features. Various studies have been conducted to analyze and combine images of different resolutions [15,16].

Furthermore, owing to the morphological similarities between DCIS and IDC, the annotations of pathologists can differ depending on their proficiency. This could lead to the deep learning model producing uncertain predictions. Therefore, handling these uncertainties is essential for obtaining better diagnostic models. To solve this problem, researchers have been working on statistical uncertainty for a long time. For example, the reject option has been applied to support vector machine, nearest neighbor, and k-nearest neighbor models [17–19]. When considering selective prediction in neural networks, the most common method is to find a rejection threshold from a pre-trained model with all the data and to train the model using that threshold [20,21]. This approach involves training using the entire dataset and determining a rejection threshold using outputs that yield ambiguous results. Subsequently, by using this rejection threshold, data are excluded during training. A drawback of this method is that it requires two rounds of training. Therefore, recent studies have devised methods to simultaneously determine the rejection threshold and exclude data based on that threshold during training [22]. In addition, methods using Monte Carlo dropout to reject uncertain data [23] and methods using softmax response are available.

Based on these studies, we have developed a model for segmenting DCIS and IDC in breast tissue. We introduce the multi-resolution selective segmentation (MurSS) model, which uses low- and high-resolution patches to leverage multi-resolution features and reject ambiguous regions using the selective segmentation method (SSM). The major contributions of this study are summarized as follows:

- We incorporate a large field of view as an explicit context within our model's architecture, improving the cancer lesion segmentation performance.
- We address the challenge of ambiguity in segmentation tasks for pathological images through the strategic implementation of selective segmentation methods.

## 2. Materials and Methods

### 2.1. Materials

For MurSS's training and evaluation, we used 1276 hematoxylin and eosin (H&E) stained WSIs. These datasets consist of 1181 TCGA [7] breast invasive carcinoma (TCGA-BRCA) and 95 Korea University Medical Center, Guro Hospital (KUMC-Guro), BRCA H&E stained WSIs. Three pathologists working at different institutions annotated the WSIs for benign, DCIS, and IDC. We created training, validation, and test datasets using majority voting for three different sets of annotated data. From this, 87 slides were excluded from the dataset due to slide quality issues that could affect the model's performance [24]. As a training dataset, 876 TCGA-BRCA WSIs were randomly selected and used, while the remaining 218 TCGA-BRCA WSIs were used for validation. The test performance was evaluated using 95 WSIs from KUMC-Guro, utilizing the model with the highest performance on the validation set. KUMC-Guro BRCA H&E stained WSIs were collected from early breast cancer patients with pN0, hormone receptor-positive, and human epidermal growth factor receptor type 2 (HER2)-negative statuses. The data used in this study were collected from January 2014 to July 2019 and scanned using an Aperio AT-2 scanner. If the patient had already undergone chemotherapy before surgery or the pathologist determined that the H&E slide condition was unsuitable for clinical histological and pathological trials, the relevant data were not collected.

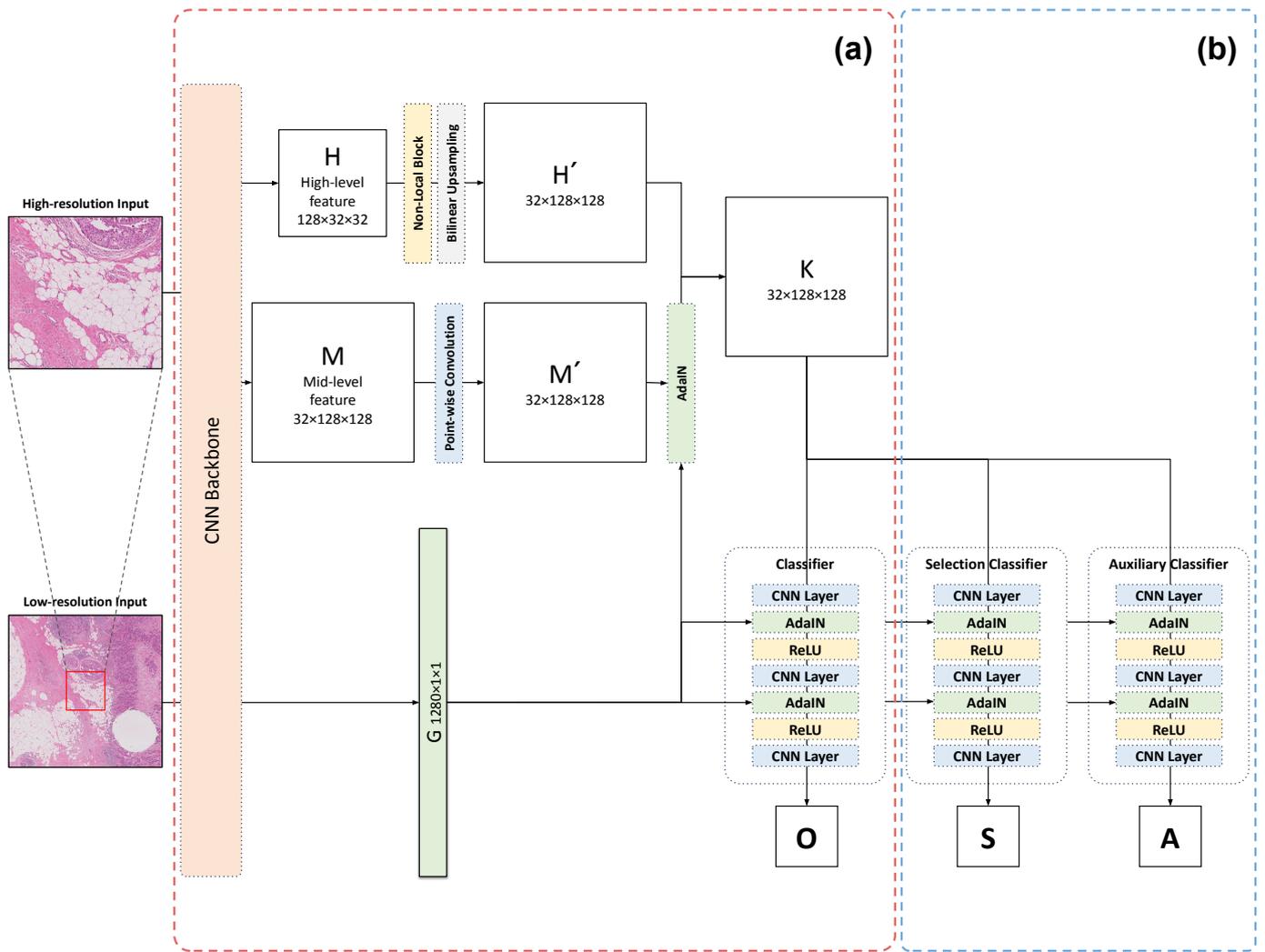
We divided each slide into  $544 \times 544$ -pixel patches for training, validation, and testing. The images were of high, mid, and low resolutions with magnifications of 5.0 $\times$ , 2.5 $\times$ , and 1.25 $\times$ , respectively. We extracted patches at a 5.0 $\times$  magnification and then used the central coordinate values to extract patches at magnification levels of 2.5 $\times$  and 1.25 $\times$ . Consequently, in the case of the 2.5 $\times$  and 1.25 $\times$  magnification patches, overlapping areas existed. Using all patches from different resolutions, we randomly cropped them to a size of  $512 \times 512$  pixels. To overcome the data imbalance problem, we oversampled the DCIS patches nine times. The same patches were utilized multiple times based on the oversampling ratio. Subsequently, hard augmentations were implemented, including flipping, rotating, blurring, cropping, color jittering, and so forth. Table S3 presents the experimental results obtained with varying oversampling ratios.

### 2.2. Methods

This section describes the key elements of the multi-resolution selective segmentation (MurSS) model. MurSS consists of the multi-resolution adaptive normalization (MurAN) model, which receives high-resolution and low-resolution images as input and produces pixel-level classification output, and the selective segmentation method (SSM), which selects the pixel-level uncertainty-based classification output. MurAN is introduced in Section 2.2.1, and the SSM is introduced in Section 2.2.2. Finally, in Section 2.2.3 we briefly explain our evaluation metrics.

#### 2.2.1. Multi-Resolution Adaptive Normalization

The architecture of our proposed model is shown in Figure 1. It processes two inputs of different resolutions, utilizing a shared-weight CNN model, termed the 'backbone', to extract multi-resolution features. The model generates mid- and high-level features through the backbone when processing high-resolution input. Conversely, for low-resolution input, after passing through the backbone, the model obtains contextual information regarding the overall style of the image.



**Figure 1.** The multi-resolution selective segmentation (MurSS) model consists of 2 parts. (a) The multi-resolution adaptive normalization (MurAN) model extracts high-level ( $H$ ) and mid-level ( $M$ ) features from high-resolution patches.  $G$  represents the extracted contextual information from low-resolution patches.  $G$  is combined with feature maps ( $K$ ) through the adaptive instance normalization (AdaIN) blocks. Finally, the classifier creates the pixel-level classification output ( $O$ ). (b) The selective segmentation method (SSM) is added to MurAN. MurSS uses three classifiers, while MuRAN uses only one classifier that produces the original output ( $O$ ).  $O$  is the original output that is produced by the prediction function,  $f$ . The selection output ( $S$ ) is produced by the selection function,  $g$ , to determine the selection coverage. The auxiliary function,  $h$ , is only used for training to generalize the loss and produce an auxiliary output ( $A$ ).

High-level, high-resolution features pass through the non-local block layer to incorporate patch-wise global features. The non-local block, also known as the self-attention mechanism, is a method proposed by [25]. Non-local blocks can capture different types of spatial information, leading to better performance [26–28]. Using the non-local block and bilinear upsampling, we obtain  $H' \in R^{B \times C \times H \times W}$ , with a size of  $B \times 32 \times 128 \times 128$ .  $B$ ,  $C$ ,  $H$ , and  $W$  denote the batch size, channel, height, and width of the feature. Mid-level features,  $M$ , have a shape of  $32 \times 128 \times 128$ , and these features go through a pointwise convolution layer and return  $M' \in R^{C \times H \times W}$  with the same shape as  $M$ . Then,  $M'$  is combined with the contextual information  $G \in R^{C \times H \times W}$  with a size of  $128 \times 1 \times 1$ . If we concatenate  $M'$  and  $G$ , the representation of each pixel may be different due to the different sizes of the features. This may cause performance degradation in the segmentation task. To solve this problem,

we used adaptive instance normalization (AdaIN), as proposed in [29]. AdaIN extends direct application in style transfer by normalizing the content of an image’s features and then applying the style image’s mean and standard deviation. This approach allows AdaIN to effectively transfer the style characteristics to the content image, maintaining its structure while adopting the style attributes. AdaIN scales  $M''$ ’s normalized content input with  $G$  and shifts it with  $G$  without any learnable affine parameters. This computation is based on batches, meaning it is carried out independently for each image within a batch. In short, AdaIN performs style transfer in the feature space by transferring feature statistics, specifically the channel-wise mean and standard deviation. Through AdaIN, we obtain  $M''$  of size  $32 \times 128 \times 128$ , represented as  $M'' \in R^{C \times H \times W}$ . Finally, the patch-wise global feature from the non-local block,  $H'$ , and  $M'$  are merged through a concatenation process. Afterward, we pass it through a classifier and produce the final output.

### 2.2.2. Selective Segmentation Method

To overcome the data uncertainties derived from the morphological similarities of DCIS and IDC, as well as the possibility of differing annotations by pathologists, we used the selective segmentation method (SSM). By using the SSM, we can automatically reject ambiguous regions. SelectiveNet [22] enables the model to reject inputs that frequently yield erroneous results, thereby mitigating the potential for uncertain predictions during both the inference and training processes. Inspired by SelectiveNet, we developed the SSM for any segmentation task. The SSM produces the output of MurAN and two additional outputs: the selection and the auxiliary output. The authors of [22] stated that the performance of a selective method can be quantified using both the coverage ratio and risk. We defined the empirical coverage risk as

$$r(I_{high}, I_{low}, y) = \frac{\sum_{i=1}^m \sum_{j=1}^n l(f(I_{high}, I_{low})_{ij}, y_{ij}) g(I_{high}, I_{low})_{ij}}{\sum_{i=1}^m \sum_{j=1}^n g(I_{high}, I_{low})_{ij}} \tag{1}$$

where  $I_{high}$  and  $I_{low}$  are the high- and low-resolution patches;  $y$  is the ground-truth label;  $l(a, b)$  is the pixel-wise cross-entropy loss;  $a$  is the output of the model; and  $b$  represents the images annotated by the pathologists. Also,  $f$  is the prediction function that produces the output and  $g$  is the selection function that produces the selection output. Equation (1) shows the empirical coverage risk regarding the selective region based on the coverage ratio. Also, the quadratic penalty term,  $L_{Penalty} = \max(0, T - \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n g(I_{high}, I_{low})_{ij})^2$ , is added, as shown in Equation (2).

$$L_{Select} = r(I_{high}, I_{low}, y) + \lambda L_{Penalty} \tag{2}$$

where  $T$  is the pre-defined value from 0 to 1 that indicates the coverage ratio the model aims to achieve and  $\lambda$  controls the relative importance of the constraints. We train MurSS to reduce the empirical selective risk to minimize  $L_{Select}$ . To determine the optimal empirical coverage risk, certain pixels must be rejected under the condition that the percentage of non-rejected pixels meets or exceeds the coverage ratio  $T$ . When  $\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n g(I_{high}, I_{low})_{ij} < T$ , the quadratic penalty term will be greater than 0, reducing  $L_{Select}$  and increasing the size of rejection regions. Therefore, there will be more opportunities to reject uncertain areas. Otherwise, when  $\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n g(I_{high}, I_{low})_{ij} \geq T$ , the quadratic penalty term will be 0, allowing the model to focus more on reducing the risk within the coverage area without reducing the size of rejection regions. However, if we calculate the loss based only on the coverage region, the model may overfit to the selected data. To address this, we use  $h(I_{high}, I_{low})$ , which represents the auxiliary prediction, to compute the auxiliary prediction risk during the training process. Therefore, our model’s loss is defined as

$$L_{Total} = \alpha L_{Select} + (1 - \alpha) L_{Aux} \tag{3}$$

where  $L_{Aux} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n l(h(I_{high}, I_{low})_{ij}, y_{ij})$ . From this,  $\alpha$  is the hyperparameter used to control the ratio of empirical risk and auxiliary prediction risk. Also, in  $L_{Aux}$ ,  $l(a, b)$  is the pixel-wise cross-entropy loss, where  $a$  is the auxiliary output of the model and  $b$  represents the images annotated by the pathologists.

### 2.2.3. Evaluation Metrics

To evaluate the performance of our proposed model, MurSS, against other segmentation models, we use two metrics: pixel-level accuracy (accuracy) and Intersection over Union (IoU). Accuracy measures the percentage of correctly classified pixels in the image (Equation (4)). The IoU is calculated as the area of the intersection over the union of the ground truth ( $S_y$ ) and prediction area ( $S_O$ ). The formulas for accuracy and the IoU are shown in Equation (5).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$IoU(S_O, S_y) = \frac{|S_O \cap S_y|}{|S_O \cup S_y|} = \frac{TP}{TP + FP + FN} \tag{5}$$

Regarding the metrics for IDC, the following confusion matrix (Table 1) shows examples of TPs, FNs, FPs, and TNs. The confusion matrix is different from binary cases.

**Table 1.** Sample confusion matrix for IDC.

Confusion Matrix for IDC		Predicted Values		
		Benign	DCIS	IDC
Pathologists	Benign	TN	FN	FP
	DCIS	FN	TN	FP
	IDC	FN	FN	TP

TN: true negative; FN: false negative; FP: false positive; TP: true positive; DCIS: ductal carcinoma in situ; IDC: invasive ductal carcinoma.

The 95% confidence intervals (CI) for accuracy and mIoU were measured using the bootstrap method. A bootstrap sample was generated via random sampling with replacement from the original test dataset, ensuring each bootstrap sample was the same size as the original test dataset. This process was independently and randomly repeated 1000 times. For each bootstrap sample, accuracy and mIoU were calculated. The bootstrap 95% confidence intervals for each of the two metrics were then estimated from these repeated measurements.

### 3. Results

This section provides a concise overview of the research, presenting the conclusions drawn from the evidence analyzed. The checkpoint obtained from the most optimal epoch of the validation dataset was used for all evaluations. The inference results on the test dataset were then measured based on these checkpoints. The performance of the validation set is shown in Table S1.

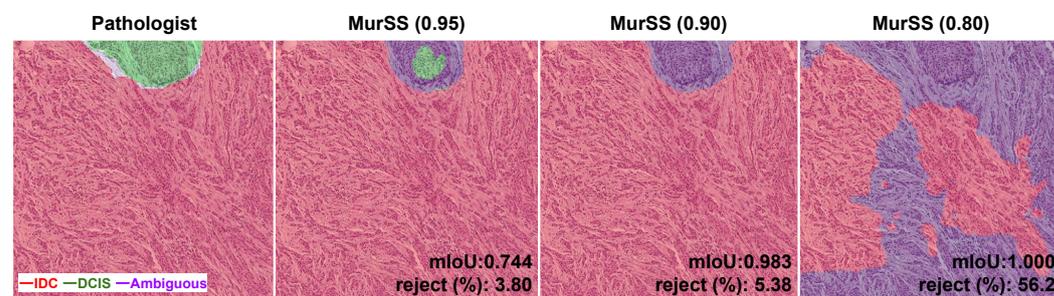
Table 2 shows a quantitative analysis of various segmentation models, including the proposed MurAN and MurSS, with predefined coverage ratios of 0.95, 0.90, and 0.80, against established models such as U-Net, HRNet, DeepLabV3, ICNet, and DMMN. The performance metrics include pixel-level accuracy (accuracy), mean Intersection over Union (mIoU), and class-specific IoUs for benign, DCIS, and IDC. To evaluate the performance of the single-resolution model, we used U-Net and compared the results obtained from high, mid, and low resolutions. The high-resolution U-Net model exhibited higher mIoU values than the other U-Net models, with increases ranging from 0.0154 to 0.0236. Based on these results, we evaluated the performance of HRNet and DeepLabV3 using the high-resolution patch with the best performance and compared it with the other multi-resolution models, MurAN and MurSS.

**Table 2.** Performance results. Comparative analysis of various segmentation models. The performance metrics include pixel-level accuracy, mean Intersection over Union (mIoU), and class-specific Intersection over Union (IoU).

Model	Coverage Ratio	Overall Measure (95% CI)		Intersection over Union (IoU)		
		Accuracy (%)	mIoU	Benign	DCIS	IDC
U-Net *	1.0	94.77 (93.39, 95.89)	0.6651 (0.6339, 0.6970)	0.9470	0.3740	0.6743
U-Net †	1.0	94.97 (93.75, 95.96)	0.6497 (0.6210, 0.6765)	0.9504	0.3360	0.6626
U-Net ‡	1.0	94.09 (92.48, 95.35)	0.6375 (0.6040, 0.6688)	0.9393	0.3463	0.6268
HRNet *	1.0	95.58 (94.57, 96.35)	0.7005 (0.6593, 0.7312)	0.9546	<b>0.4361</b>	0.7106
DeepLabV3 *	1.0	95.60 (94.71, 96.35)	0.7013 (0.6631, 0.7341)	0.9555	0.4360	0.7123
ICNet *†‡	1.0	94.55 (93.49, 95.40)	0.6714 (0.6349, 0.7016)	0.9437	0.4008	0.6698
DMMN *†‡	1.0	94.42 (93.23, 95.36)	0.6424 (0.6170, 0.6699)	0.9436	0.3287	0.6549
MurAN *†	1.0	95.88 (94.85, 96.71)	0.7055 (0.6640, 0.7399)	0.9577	0.4260	0.7328
<b>MurSS *†</b>	<b>0.95</b>	<b>96.88</b> <b>(95.97, 97.62)</b>	<b>0.7283</b> <b>(0.6865, 0.7640)</b>	<b>0.9690</b>	0.4324	<b>0.7833</b>
MurSS *†	0.90	97.09 (96.13, 97.85)	0.7356 (0.6970, 0.7705)	0.9707	0.4363	0.7999
<i>MurSS *†</i>	<i>0.80</i>	<i>98.30</i> <i>(97.53, 98.86)</i>	<i>0.7603</i> <i>(0.7067, 0.8061)</i>	<i>0.9839</i>	<i>0.4485</i>	<i>0.8487</i>

\*: high-resolution input (5.0×); †: mid-resolution input (2.5×); ‡: low-resolution input (1.25×). **Bold:** The model utilized throughout the entirety of the manuscript. *Italic:* The state-of-the-art performance.

The performance of MurSS improved, indicating its ability to automatically classify uncertain areas that may be difficult or ambiguous for pathologists. The best numerical performance was achieved with a coverage ratio of 0.80 (MurSS<sub>0.80</sub>). However, this may not provide a reliable diagnosis, as the measurement was taken with a coverage ratio of 0.80, indicating that 20% of regions were ambiguous. In the visualization presented in Figure 2, the purple regions indicate that MurSS<sub>0.80</sub> rejected too many regions. MurSS<sub>0.80</sub> rejected a significant amount of DCIS and IDC. The trained MurSS with a coverage ratio of 0.95 (MurSS<sub>0.95</sub>) accurately identified the obvious areas while only rejecting some of the ambiguous regions. Therefore, we used MurSS<sub>0.95</sub> for our qualitative analysis.

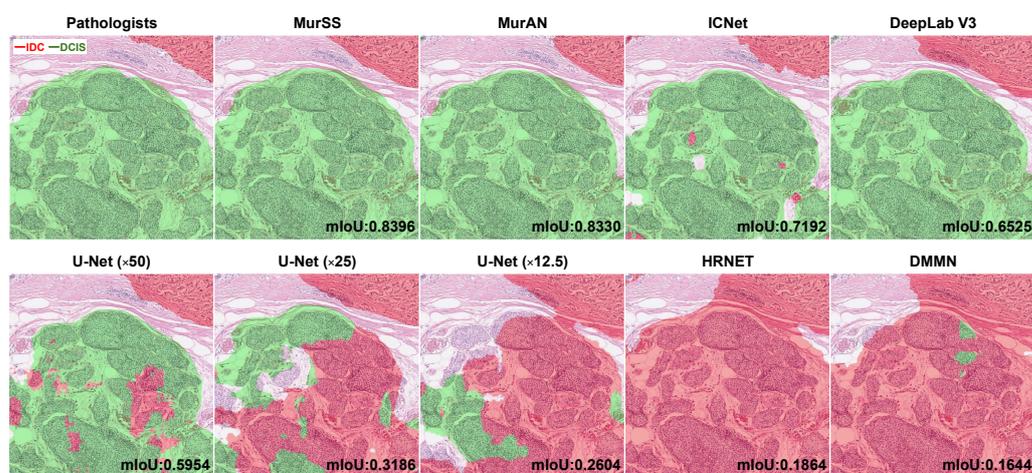


**Figure 2.** Visualization results of MurSS. Comparison between pathologists and MurSS with different coverage ratios. Green areas represent DCIS, red areas represent IDC, and purple areas represent the rejected regions. MurSS with a coverage ratio of 0.80 shows the highest mIoU, but it rejects 56.2% of regions. The visualization results show it is unreliable for diagnosis.

The proposed MurSS<sub>0.95</sub> achieved the highest performance compared to other models by rejecting 5% of WSI regions, resulting in an mIoU of 0.7283 and an accuracy of 96.88%.

Additionally, the proposed MurAN model achieved an mIoU of 0.7055 and an accuracy of 95.88%, which was the second-highest overall measure after MurSS<sub>0.95</sub>. When examining the class-specific IoU, MurSS<sub>0.95</sub> and MurAN showed the best performance for benign and IDC, while HRNet and DeepLabV3 showed the best performance for DCIS. Also, a statistical analysis was conducted to verify MurSS<sub>0.95</sub>'s superiority. The results of the analysis confirm the superiority of MurSS<sub>0.95</sub> and MurAN. Detailed results of our statistical analysis are presented in Table S2.

Figure 3 shows the visualization result and mIoU of each patch with multiple deep learning models. The green areas indicate the regions where the models predicted DCIS, and the red areas indicate the regions where models predicted IDC. Generally, MurSS<sub>0.95</sub>, MurAN, DeepLabV3, and HRNet performed well regarding mIoU. However, MurAN, DeepLabV3, and HRNet exhibited overconfidence in their results, leading to misjudgments of IDC and DCIS. Further visualization results are presented in Figure S1a–c.



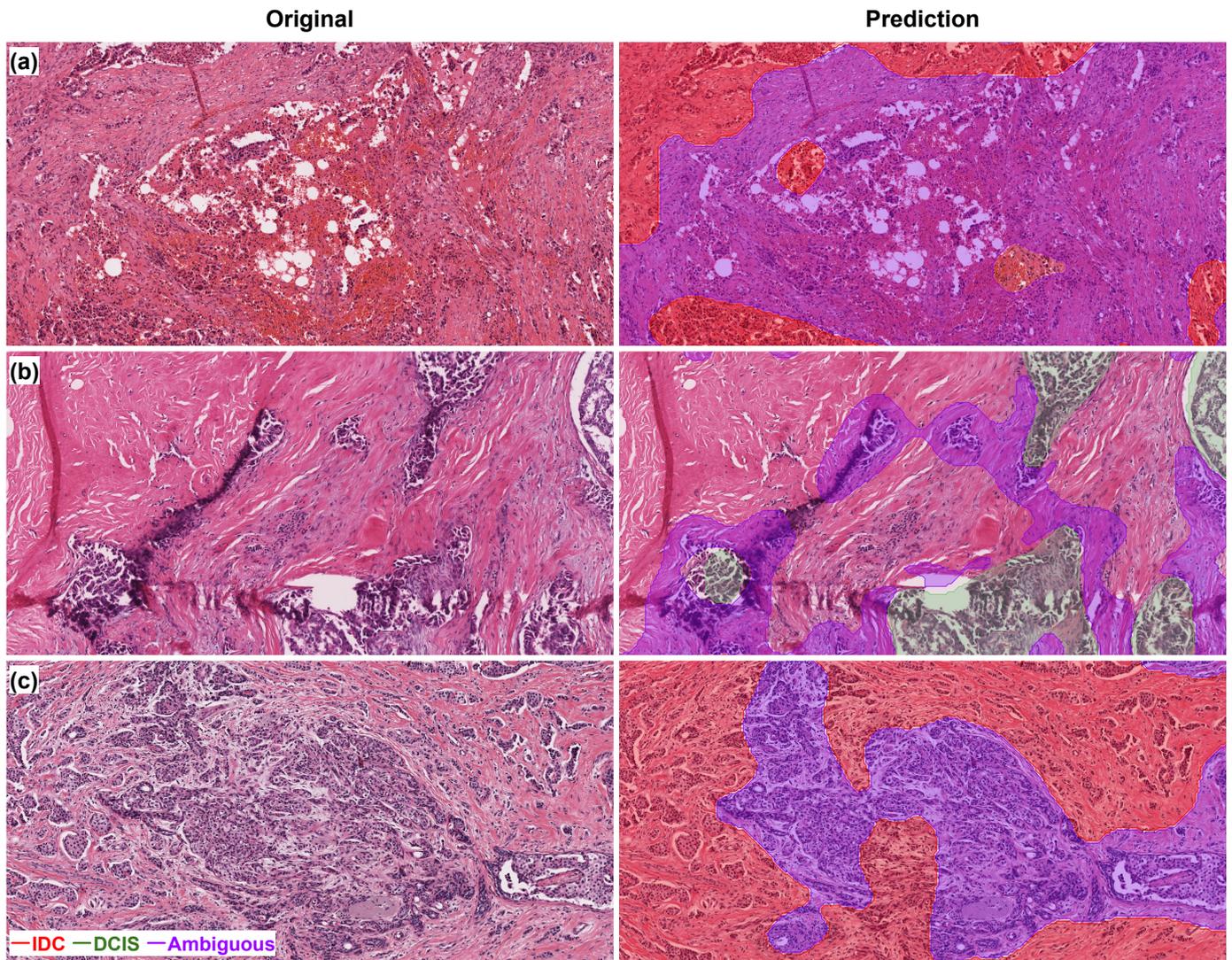
**Figure 3.** Visualization results. Comparison between pathologists and deep learning models. The mIoU was measured at the patch level. Green areas represent DCIS, and red areas represent IDC.

#### 4. Discussion

This study proposes a segmentation model for accurate breast cancer diagnosis. MurSS attempts to overcome the limitations of existing methods by combining contextual and content information from various resolutions using adaptive instance normalization. Additionally, it effectively performs segmentation for uncertain areas using the SSM. To measure the uncertainty, the Intersection over Union (IoU) from DCIS annotations by two pathologists was calculated. The slide-level IoU of DCIS was 0.44 between the two pathologists, showing that DCIS can be interpreted differently.

##### 4.1. Visualization Explanation

The visualization results are provided in Figure 4. MurSS identified four different classes (benign, DCIS, IDC, and ambiguous) with different colors on the slide. According to the model, the red region was IDC, the green region was DCIS, and the purple region was ambiguous. Purple regions usually surround the red and green regions somewhat since they represent the border of benign, IDC, or DCIS. When we focused on the regions with a high volume of purple, we discovered that these regions exhibited ambiguous histomorphological characteristics. As shown in Figure 4, MurSS identified regions with tissue artifacts as ambiguous (Figure 4a). In these areas, tumor cell nests were detected without a clear myoepithelial cell layer (Figure 4b). Additionally, the model diagnosed the lesion as vague in areas where sclerosing adenosis and small tumor cell nests were intermingled (Figure 4c).



**Figure 4.** Visualization results of MurSS. The model identified regions with tissue artifacts as ambiguous (a). The ambiguous region also showed tumor cell nests without a definite myoepithelial cell layer (b). The model also identified lesions in the area where sclerosing adenosis and small tumor cell nests were intermingled as vague (c). Slide regions were classified as benign (no color), IDC (red), DCIS (green), and ambiguous (purple).

#### 4.2. Limitations and Future Work

Despite the use of MurSS, the performance regarding DCIS remains low. This low performance is likely due to the minimal amount of DCIS data available for training, accounting for only 0.5% of all pixels. This data imbalance can negatively impact the performance of deep learning models [30]. We conducted multiple experiments, such as oversampling the DCIS, weighted cross-entropy loss, etc. The results are presented in Tables S3–S5. Collecting a significant amount of DCIS data is necessary. Additionally, our model demonstrated a gap in performance between the validation and test sets. Differences in the intensity of data scanning and staining for each institution may account for the observed variations. Studies have shown that the staining technique significantly affects performance [31]. A normalization method for data collected from various institutions is needed to address this issue. The normalization method still presents a challenge for future work. MurSS and other deep learning models require significant improvement to be applicable in medical image analysis. A multi-resolution approach based on pathology data understanding, rather than solely learning with a single resolution, is a promising strategy.

Also, in medical image analysis, where uncertain results or data can have significant consequences, it is advisable to acknowledge and address these uncertainties. Our proposed solution, SSM, offers one possible approach to address these challenges. It also leverages the accurate segmentation of invasive areas to automate tumor type and histologic grade analyses in future research.

## 5. Conclusions

This paper presents MurSS to address the challenge of training deep learning models with gigapixel-sized, high-resolution WSIs. To reduce computational cost, each WSI is divided into smaller patches. However, this approach may cause the model to focus solely on local information rather than the entire WSI, potentially leading to reduced performance. To address these issues, MurSS overcomes the problem of relying solely on local information by using both local content information, obtained at a high resolution, and global contextual information, obtained at a low resolution.

Additionally, for stable training and accurate diagnosis, deep learning models require the accurate labeling of morphologically similar but distinct conditions, such as DCIS and IDC. This histomorphological similarity can result in inconsistent annotations, even among expert pathologists. To address this, SSM is proposed to improve performance by rejecting inaccurate annotations caused by morphological similarities. Therefore, MurSS can identify ambiguous regions and allow pathologists to review them. The results show that MurSS achieved an improvement in mIoU ranging from 0.0717 to 0.1234 compared to existing models. MurSS also achieved a pixel-level accuracy of 98.28%, which is a 2.61% to 4.25% improvement over previous models. Based on these findings, we propose research avenues for addressing the limitations of medical imaging.

**Supplementary Materials:** The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/bioengineering11050463/s1>, Table S1: Validation Result from the Best Epoch; Table S2: Test Result Statistical Analysis; Table S3: Sampled Test about Oversampling DCIS; Table S4: Sampled Test about Weighted Cross Entropy Loss; Table S5: Sampled Test about Data Uncertainty; Figure S1a–c: Visualization Results.

**Author Contributions:** Conceptualization, J.L., G.L. and T.-Y.K.; methodology, J.L.; software, J.L.; validation, J.L., M.-S.J., C.K. and H.C.; formal analysis, J.L.; investigation, J.L.; resources, J.L.; data curation, J.L.; writing—original draft preparation, J.L.; writing—review and editing, T.-Y.K., M.-S.J., C.K. and H.C.; visualization, J.L.; supervision, T.-Y.K. and H.C.; project administration, T.-Y.K. and H.C.; funding acquisition, S.W.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** This study was reviewed and approved by the Institutional Review Board (IRB) of Korea University Medical Center, Seoul, Korea (IRB No. 2020-GRO-146).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The code used in this article cannot be published due to privacy but can be obtained from the corresponding author upon reasonable request.

**Conflicts of Interest:** Joonho Lee, Geongyu Lee, and Hyeyoon Chang are employees of Deep Bio Inc. Tae-Yeong Kwak is the CTO of Deep Bio Inc., and Sun Woo Kim is the CEO of Deep Bio Inc. The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

Accuracy	pixel-level accuracy (%)
AdaIN	adaptive instance normalization
BRCA	breast invasive carcinoma
CI	confidence interval
CNN	convolutional neural network
DCIS	ductal carcinoma in situ
H&E	hematoxylin and eosin
IDC	invasive ductal carcinoma
IoU	Intersection over Union
KUMC-Guro	Korea University Medical Center, Guro Hospital
mIoU	mean Intersection over Union
MurAN	multi-resolution adaptive normalization
MurSS	multi-resolution selective segmentation
SSM	selective segmentation method
TCGA	The Cancer Genome Atlas
WSI	whole-slide image

## References

1. Chhikara, B.S.; Parang, K. Global Cancer Statistics 2022: The trends projection analysis. *Chem. Biol. Lett.* **2023**, *10*, 451.
2. Hophan, S.L.; Odnokoz, O.; Liu, H.; Luo, Y.; Khan, S.; Gradishar, W.; Zhou, Z.; Badve, S.; Torres, M.A.; Wan, Y. Ductal carcinoma in situ of breast: From molecular etiology to therapeutic management. *Endocrinology* **2022**, *163*, bqac027. [[CrossRef](#)] [[PubMed](#)]
3. Dillon, D.; Guidi, A.; Schnitt, S. Pathology of invasive breast cancer. *Dis. Breast* **2010**, *5*, 381–410.
4. Boyages, J.; Recht, A.; Connolly, J.L.; Schnitt, S.J.; Gelman, R.; Kooy, H.; Love, S.; Osteen, R.T.; Cady, B.; Silver, B.; et al. Early breast cancer: Predictors of breast recurrence for patients treated with conservative surgery and radiation therapy. *Radiother. Oncol.* **1990**, *19*, 29–41. [[CrossRef](#)] [[PubMed](#)]
5. Corsi, F.; Albasini, S.; Ciciriello, S.; Villani, L.; Truffi, M.; Sevieri, M.; Sorrentino, L. Extensive Intraductal Component in Breast Cancer: What Role in Disease-Free Survival? *J. Surg. Res.* **2023**, *283*, 233–240. [[CrossRef](#)]
6. Elling, D.; Vesper, A.; Fiedler, B.; Martin, H.; Krockner, J. Intraductal component in invasive breast cancer: Analysis of 250 resected surgical specimens. *Breast* **2001**, *10*, 405–410. [[CrossRef](#)]
7. Tomczak, K.; Czerwińska, P.; Wiznerowicz, M. Review The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge. *Contemp. Oncol. Onkol.* **2015**, *2015*, 68–77. [[CrossRef](#)]
8. Luo, W.; Li, Y.; Urtasun, R.; Zemel, R. Understanding the effective receptive field in deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2016**, *29*, 2–5.
9. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Proceedings, Part III 18; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
10. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
11. Sun, K.; Xiao, B.; Liu, D.; Wang, J. Deep high-resolution representation learning for human pose estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 5693–5703.
12. Miangoleh, S.M.H.; Dille, S.; Mai, L.; Paris, S.; Aksoy, Y. Boosting monocular depth estimation models to high-resolution via content-adaptive multi-resolution merging. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 9685–9694.
13. Zhao, H.; Qi, X.; Shen, X.; Shi, J.; Jia, J. Icnet for real-time semantic segmentation on high-resolution images. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 405–420.
14. Ho, D.J.; Yarlagadda, D.V.; D’Alfonso, T.M.; Hanna, M.G.; Grabenstetter, A.; Ntiemoah, P.; Brogi, E.; Tan, L.K.; Fuchs, T.J. Deep multi-magnification networks for multi-class breast cancer image segmentation. *Comput. Med Imaging Graph.* **2021**, *88*, 101866. [[CrossRef](#)]
15. Alsubaie, N.; Shaban, M.; Snead, D.; Khurram, A.; Rajpoot, N. A multi-resolution deep learning framework for lung adenocarcinoma growth pattern classification. In Proceedings of the Medical Image Understanding and Analysis: 22nd Conference, MIUA 2018, Southampton, UK, 9–11 July 2018; Proceedings 22; Springer: Berlin/Heidelberg, Germany, 2018; pp. 3–11.

16. Sirinukunwattana, K.; Alham, N.K.; Verrill, C.; Rittscher, J. Improving whole slide segmentation through visual context—A systematic study. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2018: 21st International Conference, Granada, Spain, 16–20 September 2018; Proceedings, Part II 11; Springer: Berlin/Heidelberg, Germany, 2018; pp. 192–200.
17. Fumera, G.; Roli, F. Support vector machines with embedded reject option. In Proceedings of the Pattern Recognition with Support Vector Machines: First International Workshop, SVM 2002, Niagara Falls, ON, Canada, 10 August 2002; Springer: Berlin/Heidelberg, Germany, 2002; pp. 68–82.
18. Hellman, M.E. The nearest neighbor classification rule with a reject option. *IEEE Trans. Syst. Sci. Cybern.* **1970**, *6*, 179–185. [[CrossRef](#)]
19. Dalitz, C. Reject options and confidence measures for knn classifiers. *Schriftenreihe Fachbereichs Elektrotechnik Inform. Hochsch. Niederrh.* **2009**, *8*, 16–38.
20. De Stefano, C.; Sansone, C.; Vento, M. To reject or not to reject: That is the question—an answer in case of neural classifiers. *IEEE Trans. Syst. Man, Cybern. Part C Appl. Rev.* **2000**, *30*, 84–94. [[CrossRef](#)]
21. El-Yaniv, R. On the Foundations of Noise-free Selective Classification. *J. Mach. Learn. Res.* **2010**, *11*, 1605–1641.
22. Geifman, Y.; El-Yaniv, R. Selectivenet: A deep neural network with an integrated reject option. In Proceedings of the International Conference on Machine Learning, PMLR, Long Beach, CA, USA, 9–15 June 2019; pp. 2151–2159.
23. Gal, Y.; Ghahramani, Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In Proceedings of the International Conference on Machine Learning, PMLR, New York, NY, USA, 19–24 June 2016; pp. 1050–1059.
24. Dodge, S.; Karam, L. Understanding how image quality affects deep neural networks. In Proceedings of the 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX), Lisbon, Portugal, 6–8 June 2016; IEEE: New York, NY, USA, 2016; pp. 1–6.
25. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-local neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7794–7803.
26. Cao, Y.; Xu, J.; Lin, S.; Wei, F.; Hu, H. Gcnet: Non-local networks meet squeeze-excitation networks and beyond. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Seoul, Republic of Korea, 27–28 October 2019.
27. Chen, Y.; Kalantidis, Y.; Li, J.; Yan, S.; Feng, J. A<sup>2</sup>-nets: Double attention networks. *Adv. Neural Inf. Process. Syst.* **2018**, *31*. [[CrossRef](#)]
28. Gao, Z.; Xie, J.; Wang, Q.; Li, P. Global second-order pooling convolutional networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3024–3033.
29. Huang, X.; Belongie, S. Arbitrary style transfer in real-time with adaptive instance normalization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 1501–1510.
30. Johnson, J.M.; Khoshgoftaar, T.M. Survey on deep learning with class imbalance. *J. Big Data* **2019**, *6*, 1–54. [[CrossRef](#)]
31. Madabhushi, A.; Lee, G. Image analysis and machine learning in digital pathology: Challenges and opportunities. *Med. Image Anal.* **2016**, *33*, 170–175. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.