

## Article

# Edge-Guided Cell Segmentation on Small Datasets Using an Attention-Enhanced U-Net Architecture

Yiheng Zhou <sup>1,2</sup>, Kainan Ma <sup>1</sup> , Qian Sun <sup>1</sup>, Zhaoyuxuan Wang <sup>1</sup> and Ming Liu <sup>1,2,\*</sup>

<sup>1</sup> Institute of Semiconductors, Chinese Academy of Sciences, Beijing 100083, China; zhouyiheng@semi.ac.cn (Y.Z.); makainan@semi.ac.cn (K.M.); sunqian@semi.ac.cn (Q.S.); wangzhaoyuxuan@semi.ac.cn (Z.W.)

<sup>2</sup> College of Materials Science and Opto-Electronic Technology, University of Chinese Academy of Sciences, Beijing 100049, China

\* Correspondence: liuming@semi.ac.cn

**Abstract:** Over the past several decades, deep neural networks have been extensively applied to medical image segmentation tasks, achieving significant success. However, the effectiveness of traditional deep segmentation networks is substantially limited by the small scale of medical datasets, a limitation directly stemming from current medical data acquisition capabilities. To this end, we introduce AttEUnet, a medical cell segmentation network enhanced by edge attention, based on the Attention U-Net architecture. It incorporates a detection branch enhanced with edge attention and a learnable fusion gate unit to improve segmentation accuracy and convergence speed on small medical datasets. The AttEUnet allows for the integration of various types of prior information into the backbone network according to different tasks, offering notable flexibility and generalization ability. This method was trained and validated on two public datasets, MoNuSeg and PanNuke. The results show that AttEUnet significantly improves segmentation performance on small medical datasets, especially in capturing edge details, with F1 scores of 0.859 and 0.888 and Intersection over Union (IoU) scores of 0.758 and 0.794 on the respective datasets, outperforming both convolutional neural networks (CNNs) and transformer-based baseline networks. Furthermore, the proposed method demonstrated a convergence speed over 10.6 times faster than that of the baseline networks. The edge attention branch proposed in this study can also be added as an independent module to other classic network structures and can integrate more attention priors based on the task at hand, offering considerable scalability.

**Keywords:** small-dataset segmentation; edge detection; medical image processing; feature fusion



**Citation:** Zhou, Y.; Ma, K.; Sun, Q.; Wang, Z.; Liu, M. Edge-Guided Cell Segmentation on Small Datasets Using an Attention-Enhanced U-Net Architecture. *Information* **2024**, *15*, 198. <https://doi.org/10.3390/info15040198>

Academic Editor: Andreas Triantafyllidis

Received: 4 March 2024  
Revised: 24 March 2024  
Accepted: 1 April 2024  
Published: 3 April 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/>).

## 1. Introduction

In the field of computer vision, medical image segmentation has emerged as a pivotal sub-discipline, drawing widespread attention from the academic community [1,2]. This technique plays a crucial role in the medical diagnostic process, especially in the qualitative labeling analysis of cell images, which significantly influences doctors' judgments of medical conditions. Traditional cell labeling analysis, reliant on individual doctors' experience, is time-consuming and susceptible to subjective biases, presenting certain limitations [3]. Against this backdrop, algorithms for automated computer analysis of medical images have become an effective alternative. Such computer-assisted cell image segmentation is crucial for enhancing doctors' efficiency and reducing misdiagnosis rates [4].

Computer segmentation methods for medical images fall into two main categories: traditional and deep learning methods. Traditional medical image segmentation methods often suffer from slow segmentation speeds and lack of versatility, failing to meet clinical demands [5]. In contrast, with the continuous advancement of deep learning segmentation algorithms, there has been a significant improvement in segmentation accuracy and efficiency. Deep learning has now become the dominant method for medical image

segmentation, with numerous deep neural networks demonstrating their superior performance [6–8].

Deep learning methods for medical image data analysis are growing rapidly. To date, convolutional neural networks (CNN) and other deep learning methods have been extensively applied in various medical image analysis tasks, providing high-performance computer-aided diagnosis (CAD) frameworks and solutions for many medical image processing sub-tasks [9,10]. From the perspective of medical image processing algorithm techniques, there are regression algorithms, object detection algorithms, and semantic segmentation algorithms. In particular, semantic segmentation offers end-to-end algorithms outputting pixel-level labels equal in size to the input images, so it is widely used in medical imaging. In terms of models used in semantic segmentation, there are CNN-based U-Net networks and their variants [11], as well as networks inspired by transformers [12].

Medical datasets are pivotal for algorithmic research based on deep learning methods. The development of any medical image segmentation algorithm is contingent on the data types of the existing medical datasets. In the field of 2D medical image segmentation, there are numerous challenges and public datasets, such as brain anatomy segmentation and gland segmentation datasets. A pivotal development in 2D medical image segmentation has been the employment of advanced neural network architectures, notably U-Net [13], UNet++ [14], RU-Net [15], Attention U-Net [16], and MedT [17] networks. Central to these developments is the U-Net model, a CNN-based structure, which has emerged as a classic framework in 2D medical image segmentation, demonstrating robust performance across various datasets. Building upon the foundational U-Net architecture, UNet++ innovatively incorporates augmented skip connections, effectively mitigating the semantic gap between feature maps. Further, RU-Net amalgamates the principles of recurrent convolutional neural networks and residual networks into the U-Net framework, thereby significantly augmenting the network's feature extraction efficacy. In a similar vein of advancement, Attention U-Net integrates attention mechanisms into the U-Net structure, which markedly enhances the model's focus on pertinent areas of the image, particularly in scenarios involving complex backgrounds. MedT networks, based on transformer mechanisms, have shown improvements of 0.06–2.19% compared to traditional U-Net networks and their variants on public datasets.

Deep learning methods typically depend on large-scale datasets, whereas medical datasets are often small. When classic deep network structures are applied to cell segmentation tasks on small datasets, these methods perform poorly in terms of accuracy and convergence speed [18,19]. Considering this, we have built upon the Attention U-Net architecture, adding an edge detection branch and fusion gate units, to propose a novel edge-attention enhanced cell nucleus segmentation network for small medical datasets, AttEUnet. Trained and validated on the MoNuSeg and PanNuke public datasets, the AttEUnet model not only effectively improves segmentation accuracy over the base network but also surpasses various baseline networks in segmentation precision. The smaller the dataset, the more pronounced the advantages of AttEUnet. Our method achieved F1 scores of 0.859 and 0.888 and Intersection over Union (IoU) scores of 0.888 and 0.794 on the MoNuSeg [20] and PanNuke [21] datasets, respectively. This method not only demonstrates higher accuracy but also indicates its propensity for easier convergence on small datasets. The rest of the paper is organized as follows: Section 2 discusses our edge-attention enhanced AttEUnet architecture. Section 3 reports experimental outcomes, and conclusions are drawn in Section 4.

## 2. Edge Attention Enhanced Medical Image Segmentation Method

### 2.1. Framework Overview

We introduce a nucleus segmentation network that leverages the Attention U-Net model as its backbone, enhanced with edge prior information encoding. This method's pipeline is divided into two pathways: a feature encoding–decoding branch and the attention enhancement branch. The feature encoding–decoding branch uses Attention

U-Net as the backbone network to progressively extract features and decode them into semantic labels. Meanwhile, the attention enhancement branch employs an edge filter to extract the image’s prior edge information and integrate it into the encoder of the backbone network. The integration is controlled by the Fusion Gate Unit (GF), a learnable parameter that adjusts the fusion ratio between the edge detection branch and the feature encoding–decoding branch, allowing for more refined control over the feature fusion process, as illustrated in Figure 1.

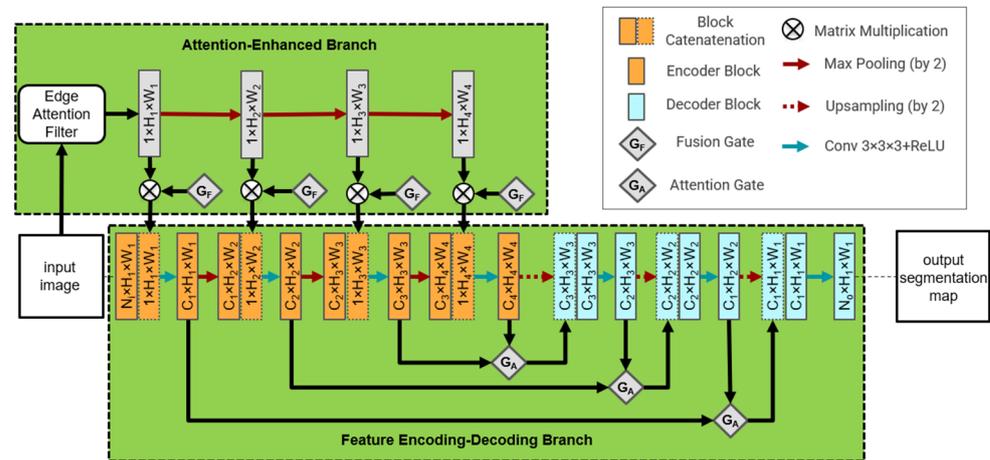


Figure 1. Pipeline of the proposed model structure.

### 2.2. Feature Encoding–Decoding Branch

Building on the foundation of Attention U-Net, we propose a more lightweight Attention U-Net encoder–decoder structure for the feature encoding–decoding branch. This branch aims to progressively encode the high-level semantic information of the original input image and decode it into semantic label maps of the same size as the input image. The encoder part specifically operates as follows: after the input image is fed into the encoder, features are extracted through filters with a  $3 \times 3$  convolution kernel size at each encoding stage, doubling the number of channels (e.g.,  $C_2 = 8C_1$ ), and downsampled by a factor of 2 using max-pooling layers (e.g.,  $H_4 = H_1/8$ ), as shown in Figure 2. During the encoding process, the outputs from various levels of the edge filtering branch are concatenated to the corresponding feature maps as an additional channel of strong prior information, contributing to the progressively deepening encoding process. The specific feature transformation process of the edge detection branch will be discussed in Section 2.3.

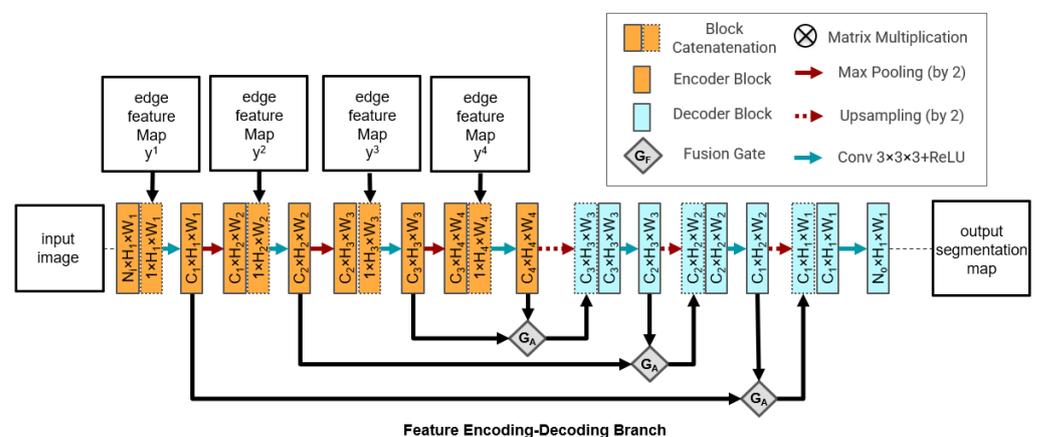


Figure 2. Lightweight Attention U-Net encoder–decoder architecture.

The decoder part of the feature encoding–decoding branch consists of up-sampling layers by a factor of 2, convolution layers, and cross-layer fusion pathways controlled

by attention gate units. The original research on Attention U-Net provides a detailed explanation of these attention gate units, which have demonstrated excellent cross-layer feature fusion capabilities across numerous datasets. Our study retains this soft attention mechanism but modifies the structure of the feature extraction layers in the encoder–decoder, reducing the number of convolution encoding blocks at each layer level to one. The addition of strong prior edge features allows for a significant reduction in network parameters while achieving more potent feature extraction capabilities.

### 2.3. Attention-Enhanced Branch

Compared to the original U-Net’s approach of directly concatenating feature maps between the encoder and decoder, the feature encoding–decoding branch based on attention mechanisms is recognized for its superior context understanding ability, beneficial for supplementing edge detail information from the encoder to the decoder. However, models based on attention mechanisms are often validated on large-scale segmentation datasets, where they can more easily learn the relationships between distant pixels. Given the smaller scale of medical imaging datasets, it is challenging to learn attention knowledge from a limited number of samples. In such cases, adding attention gates to the existing backbone network may decrease accuracy, manifesting as rough segmentation edges and internal voids in cell segmentation datasets.

To address this issue, we propose an innovative approach to incorporate artificial attention priors into the backbone network’s pipeline, termed the Attention-Enhanced Branch. This branch allows for the customization of prior types based on specific tasks. In the cell segmentation domain addressed in this paper, where the primary task involves cell identification, the most pronounced feature of cells is the morphology of their edge membranes. Therefore, we chose an edge filter to extract features of the cell membrane and integrated it as a strong prior into the feature extraction network. Considering computational complexity and algorithm interpretability, we used a traditional image edge detection filter as the edge attention enhancement module. The modulation is performed through a learnable Fusion Gate, which controls the proportion of edge prior information integrated into the backbone network. The structure of the edge detection branch is shown in Figure 3.

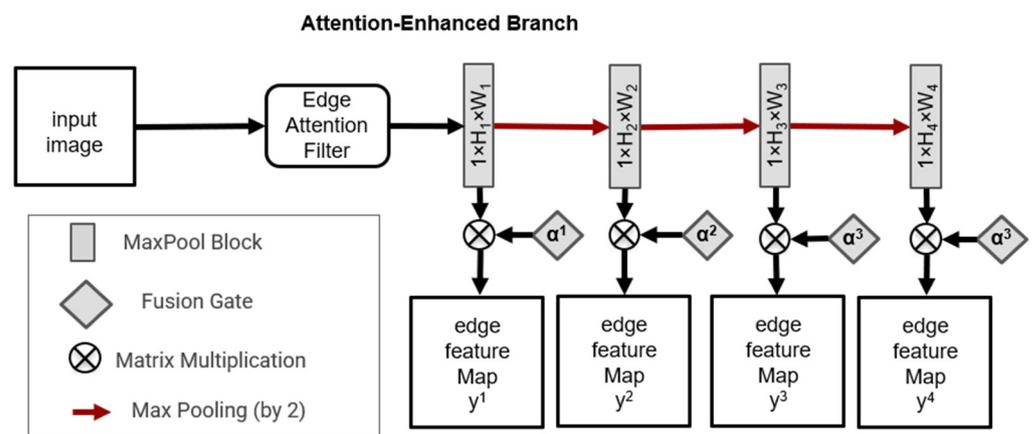


Figure 3. Network structure of the edge detection module and the Fusion Gate unit.

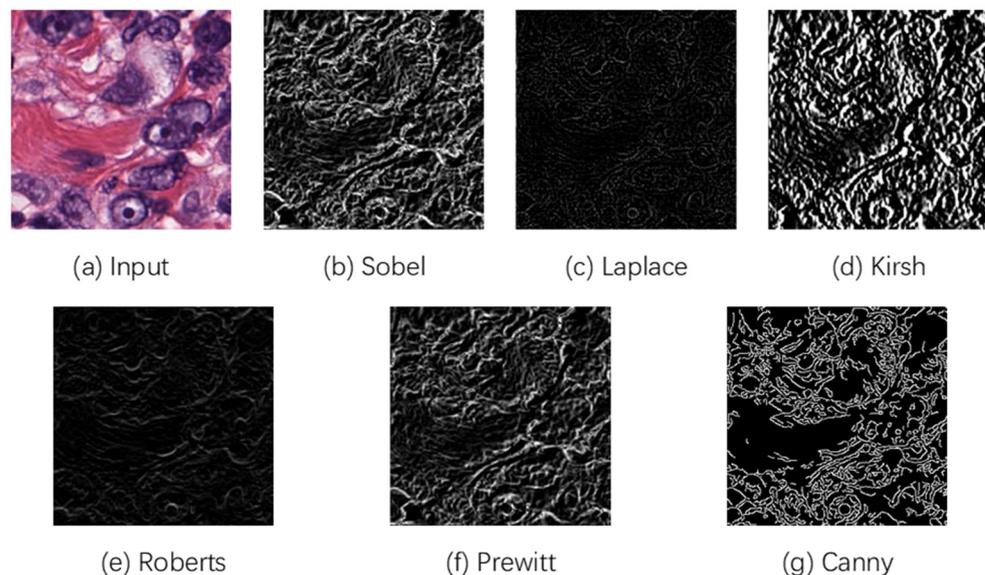
In practical applications, the selection of the edge filtering operator must be tailored to the specific requirements of the application. In the cell datasets studied in this article, cell samples are stained to appear pink within the field of view of the images, where the foreground of cells and the background have minor differences in the RGB space. However, irrelevant elements such as cell fragments are more prominent in the image compared to the background. Thus, a filtering operator with a higher tolerance for noise should be employed for medical imaging tasks. The filtering effects of various operators on the cell segmentation

dataset are shown in Figure 4. Single-stage operators like Sobel, Laplacian, and Roberts are more sensitive to noise, resulting in poorer performance on medical imaging datasets. We chose the Canny filter for edge detection, highlighting its effectiveness in cell segmentation datasets as demonstrated in Figure 4. The Canny filtering algorithm is known for its robustness, high precision, mature methodology, and low complexity. Its process involves four steps: smoothing the image with a Gaussian filter to reduce noise, detecting image edges by calculating gradient directions using the Sobel operator, removing rough edges through non-maximum suppression while preserving fine edges, and reducing noise effects by using double threshold edge detection. Subsequently, the original image, after being processed by the Canny filter and three stages of  $2 \times$  max-pooling, yields four edge-prior feature maps of different sizes, corresponding to the four feature levels in the encoder of the feature encoding–decoding branch. The edge-informed feature maps, weighted by the Fusion Gate unit, are then concatenated to the respective feature maps, integrating into the feature encoding–decoding process. The formulaic expression of the feature map integration process in the edge detection branch is as follows:

$$x_{i,j}^{k+1} = \max_{h=0}^{H-1} \max_{w=0}^{W-1} x_{i,j}^k, k = 1, 2, 3, \quad (1)$$

$$y^k = \alpha^k \cdot x^k, k = 1, 2, 3, \quad (2)$$

where  $x^k \in \mathbb{R}^{1 \times \frac{H}{2^k} \times \frac{W}{2^k}}$  represents the edge feature maps processed by max-pooling from the original Canny edge features, and  $y^k \in \mathbb{R}^{1 \times \frac{H}{2^k} \times \frac{W}{2^k}}$  denotes the Fusion Gate-modulated feature map, subsequently integrated into the backbone feature extraction network.  $x_{i,j}^k$  and  $y_{i,j}^k$  are the pixel values at position  $(i, j)$  in the two feature maps, respectively.



**Figure 4.** The performance of various filtering operators on a cell segmentation dataset.

### 3. Experiments and Results

#### 3.1. Dataset Details

We evaluated our proposed method using the MoNuSeg and PanNuke datasets, both of which are extensively used in medical cell segmentation. Unlike other fields in artificial intelligence, medical image datasets are generally small due to strong privacy concerns and the difficulty of data collection, with quantities ranging from a few dozen to a few thousand image samples. As indicated in Table 1, widely used datasets in popular AI domains contain vastly larger sample sizes. The MoNuSeg dataset, with 42 samples, is considered small in the domain of medical image processing. Conversely, the PanNuke dataset, containing

2656 samples, is regarded as medium to large. We deliberately chose one larger and one smaller dataset from the medical imaging field to test the generalizability of our approach, making the MoNuSeg and PanNuke datasets particularly representative.

**Table 1.** Several widely used datasets in popular domains of artificial intelligence.

| Domain                      | Dataset            | Sample Size |
|-----------------------------|--------------------|-------------|
| Medical Datasets            | DRIVE              | 40          |
|                             | MoNuSeg            | 42          |
|                             | LIDC-IDRI          | 1018        |
|                             | PanNuke            | 2656        |
| Autonomous Driving Datasets | Waymo Open Dataset | 200,000     |
|                             | Cityscapes         | 20,000      |
| Object Detection Datasets   | ImageNet           | 14,000,000  |
|                             | COCO               | 200,000     |
|                             | PASCAL VOC         | 10,000      |

The MoNuSeg dataset was acquired by carefully annotating tissue images from patients diagnosed with various organ tumors across multiple hospitals. It was created by downloading H&E stained tissue images captured at a  $40\times$  magnification from the TCGA archives. PanNuke is a dataset for nucleus instance segmentation and classification, featuring exhaustive nucleus annotations across 19 different tissue types, totaling 205,343 labeled nuclei, each with an instance segmentation mask. To meet our needs for cell semantic segmentation, we specifically selected cancer cells and lymphocytes as positive foreground examples, with the remainder serving as the background.

For both public datasets, we applied a ten-fold data augmentation, probabilistically performing horizontal and vertical flips, rotations, adding moderate occlusions and noise, and altering brightness and contrast. Additionally, RGB channel values of image pixels were normalized from  $[0, 255]$  to  $[0, 1]$ . The output feature maps from our proposed edge detection module will also have pixel values between  $[0, 1]$  to comply with layer normalization requirements.

### 3.2. Implementation Details

We employed the binary cross-entropy (BCE) loss between the predicted and ground truth labels for training our network, expressed as follows:

$$L_{CE(p,\hat{p})} = -\left(\frac{1}{wh} \sum_{x=0}^{w-1} \sum_{y=0}^{h-1} (p(x,y) \log(\hat{p}(x,y)) + (1-p(x,y))(1-\log(\hat{p}(x,y))))\right) \quad (3)$$

where  $w$  and  $h$  represent the image's width and height,  $p(x,y)$  corresponds to the pixel's value in the image, and  $\hat{p}(x,y)$  denotes the model's output prediction at location  $(x,y)$ . Our experiments were conducted on four NVIDIA GeForce RTX 3090 graphics cards, utilizing the PyTorch deep learning framework. The batch size was set to 40, with an Adam optimizer and a learning rate of 0.001, over 200 epochs.

To validate the superiority of our proposed network, we compared it against baselines using both CNN and transformer architectures. CNN baselines included U-Net, UNet++, and Attention U-Net, while MedT served as the transformer baseline. To evaluate the effectiveness of our Attention-Enhanced branch and the fusion gate  $GF$ , we added a Canny edge feature channel to the input of the Attention U-Net model without including the Attention-Enhanced branch and fusion module (referred to as Attention U-Net\* for baseline comparison).

### 3.3. Metrics

For quantitative analysis, we utilized F1 and IoU scores as accuracy metrics, alongside Training Time per Epoch, Inference Time, and First Achievement Time for a 0.5 IoU Score to evaluate the computational demand and convergence speed of the model. The F1 score, which calculates the harmonic mean of precision and recall, is particularly useful for assessing imbalanced datasets. An F1 score ranges from 0 to 1, with 1 indicating perfect precision and recall and 0 the worst performance. The IoU is widely used in object detection, instance segmentation, and semantic segmentation tasks as a measure of the overlap between the predicted and actual bounding boxes. IoU values also range from 0 to 1, with higher values indicating better alignment between the predicted and actual bounding boxes. Under identical hardware conditions, we measured each model's Training Time per Epoch and Inference Time across multiple trials to calculate an average, reflecting the model's computational load. Reaching a 50% IoU score for the cell foreground was considered an initial completion of the cell segmentation task. Thus, we repeatedly measured the First Achievement Time for a 0.5 IoU Score and calculated an average to serve as an indicator of the time it takes for the model to complete the cell segmentation task.

For qualitative analysis, we plotted Precision-Recall (PR) curves, Gradient-weighted Class Activation Mapping (Grad-CAM) images, and predictions for challenging samples to visualize and analyze the prediction results of different models. PR curves evaluate the comprehensive performance of classifiers under class imbalance by plotting the relationship between precision and recall at different thresholds. Grad-CAM provides visual explanations by highlighting the most crucial parts of the image for making a specific classification decision. To demonstrate the fusion capability of AttEUnet with edge prior feature maps, we visualized the first convolutional layer of various networks.

### 3.4. Results

For quantitative analysis, we evaluated the networks using F1 and IoU scores, with the results shown in Table 2. When comparing AttEUnet with baseline networks across two different datasets, it is evident that for the MoNuSeg dataset, which has fewer samples, our proposed AttEUnet network achieved an F1 improvement ranging from 0.027 to 0.176 and an IoU improvement from 0.019 to 0.18 over the baseline networks. For the PanNuke dataset, which has a slightly larger sample size, AttEUnet only showed an F1 improvement of 0.029 to 0.112 and an IoU improvement of 0.133 to 0.138, not as significant as its performance on the MoNuSeg dataset. This indicates that our proposed method, by integrating edge prior features into the backbone network, performs better on smaller datasets, effectively addressing the challenge of learning small object features due to the scarcity of samples in medical small datasets. Moreover, compared to the original Attention U-Net network, our method showed an improvement of 0.027 and 0.026 in F1 and IoU scores on the MoNuSeg dataset, and 0.044 and 0.053 on the PanNuke dataset, respectively. This suggests that the integration of robust edge prior information into the feature extraction branch not only enhances the performance of the Attention U-Net network structure itself but also surpasses other baseline networks.

We utilized Training Time per Epoch, Inference Time, and First Achievement Time for a 0.5 IoU Score as metrics for assessing the computational load and convergence speed of our model, as illustrated in Table 3. Despite adding an Attention-Enhanced Branch to the original Attention U-Net in our proposed AttEUnet, it required only 9.7% of the original network's First Achievement Time for a 0.5 IoU Score. This not only avoided adding excessive hardware training burdens but also accelerated the network's convergence speed.

**Table 2.** Quantitative analysis of AttEUnet compared to various baseline networks based on F1 and IoU scores.

| Type                     | Network                       | F1 (MoNuSeg) | IoU (MoNuSeg) | F1 (PanNuke) | IoU (PanNuke) |
|--------------------------|-------------------------------|--------------|---------------|--------------|---------------|
| CNN<br>baselines         | U-Net                         | 0.820        | 0.718         | 0.859        | 0.761         |
|                          | UNet++                        | 0.817        | 0.714         | 0.858        | 0.760         |
|                          | Attention U-Net               | 0.832        | 0.732         | 0.844        | 0.741         |
|                          | Attention U-Net* <sup>1</sup> | 0.837        | 0.739         | 0.852        | 0.751         |
| Transformer<br>baselines | MedT                          | 0.683        | 0.578         | 0.776        | 0.656         |
| Proposed                 | AttEUnet                      | 0.859        | 0.758         | 0.888        | 0.794         |

<sup>1</sup> Attention U-Net model without including the Attention-Enhanced branch and fusion module.

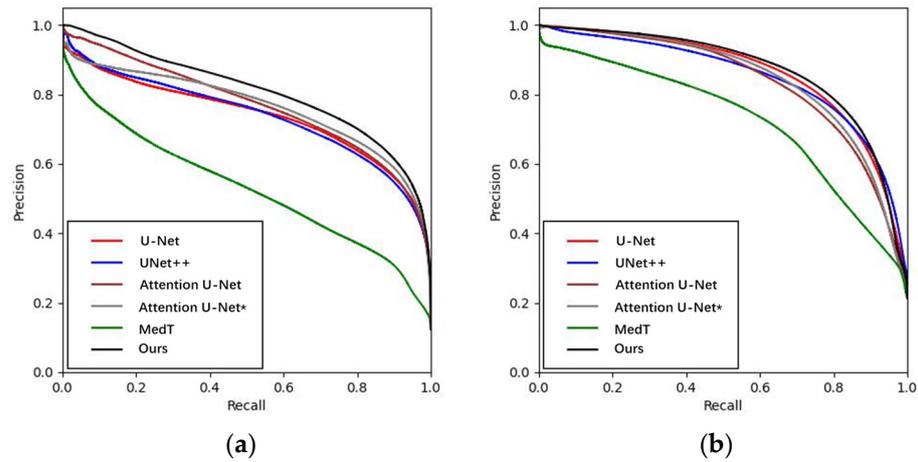
**Table 3.** Comparison of training and inference durations between AttEUnet and baseline networks.

| Network                       | Training Time per Epoch (s) | Inference Time (s) | First Achievement Time for a 0.5 IoU Score (min) |
|-------------------------------|-----------------------------|--------------------|--|
| U-Net                         | 210                         | 9.42               | 606  |
| UNet++                        | 523                         | 10.51              | 1824   |
| Attention U-Net               | 218                         | 9.72               | 579  |
| Attention U-Net* <sup>1</sup> | 275                         | 11.45              | 1039   |
| MedT                          | 1073                        | 257.08             | -  |
| AttEUnet                      | 231                         | 11.32              | 56   |

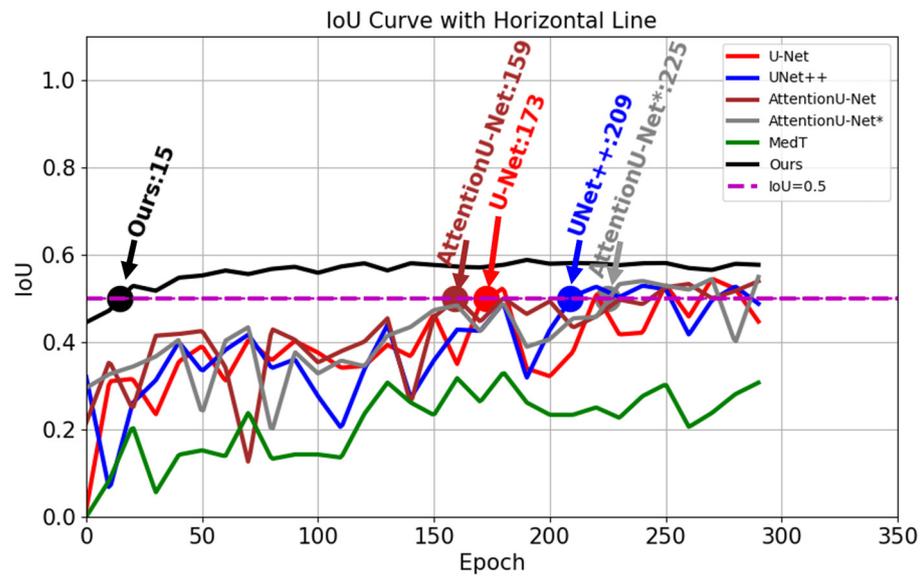
<sup>1</sup> Attention U-Net model without including the Attention-Enhanced branch and fusion module.

The First Achievement Time for a 0.5 IoU Score of the Attention U-Net\* network was 1.79 times longer than that of the original Attention U-Net and 18.55 times that of our proposed AttEUnet, paradoxically reducing the network's performance. This demonstrates that simply incorporating traditional filtering algorithms into deep networks does not necessarily enhance their effectiveness. It indirectly validates the efficacy of the Attention-Enhanced Branch and Fusion Gate pipeline introduced in this paper.

For qualitative analysis, we plotted PR curves of AttEUnet and various baseline networks on the MoNuSeg and PanNuke datasets, the number of epochs required to achieve a foreground IoU of 0.5, and visualized the predictions and Grad-CAM images of challenging samples. As seen in PR curves from Figure 5a,b, the MedT network struggles to leverage its transformer structure advantage on smaller sample sets, while our proposed AttEUnet's PR curves outperform those of all baseline networks on both datasets. Figure 5a shows the PR curve for the smaller MoNuSeg dataset, highlighting our method's more pronounced accuracy advantage over baseline networks when sample sizes are limited. Typically, achieving a foreground IoU of 0.5 is considered a basic functionality of the network; the IoU curves for segmentation of the foreground on the MoNuSeg dataset by various networks are shown in Figure 6. Benefiting from the edge attention's prior enhancement, our network required only 15 epochs to reach a foreground segmentation IoU of 0.5. In contrast, baseline networks needed 159 to 225 epochs, making our proposed network's convergence speed 10.6 times faster than that of the baselines.



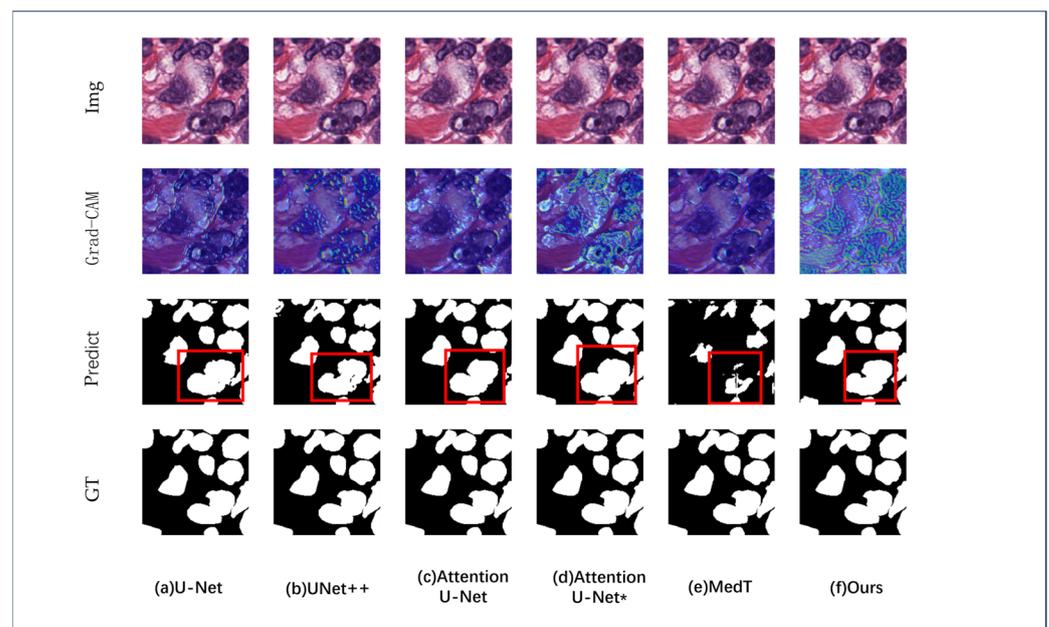
**Figure 5.** PR curves of AttEUNet and various baseline networks on two datasets: (a) MoNuSeg dataset; (b) PanNuke dataset. Note: “Attention U-Net\*” denotes the Attention U-Net model variant without the Attention-Enhanced branch and fusion module.



**Figure 6.** Epochs required for each network to first achieve a foreground IoU of 0.5 on the MoNuSeg dataset. Note: “Attention U-Net\*” denotes the Attention U-Net model variant without the Attention-Enhanced branch and fusion module.

Furthermore, designing effective network structures to address real-world clinical problems and provide interpretable outputs is crucial. Deep learning networks are often labeled as uninterpretable, posing significant challenges for clinical practice. To enhance the interpretability of our proposed AttEUNet network, we first selected a CNN-based backbone network. The local connectivity and parameter sharing characteristics of CNNs equip each convolutional kernel with strong interpretability for recognizing specific patterns, allowing us to infer the particular shapes or structures learned by the kernels. Secondly, we incorporated traditionally interpretable edge filtering feature maps as artificial attention priors into the backbone network. Traditional edge filtering algorithms, usually based on clear gradients and derivatives without involving extensive parameters or deep network computations, offer strong intuitiveness and high computational efficiency. This approach improves accuracy and workflow while ensuring the interpretability of computer-aided diagnosis tools. We utilized Grad-CAM images to highlight the target objects recognized by the neural network.

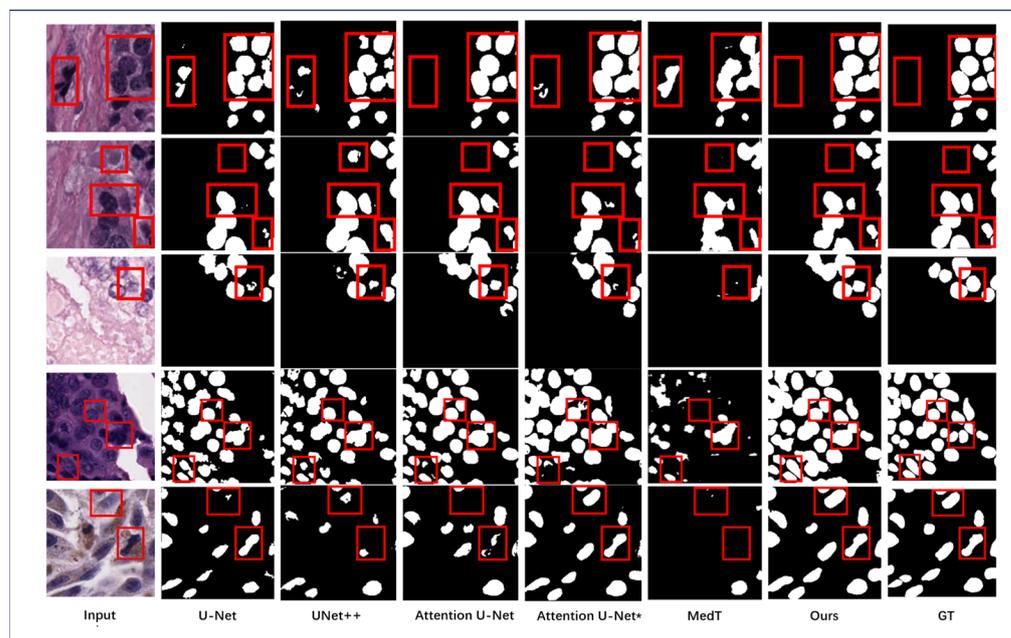
The prediction results and visualizations of challenging samples from both datasets are shown in Figure 7. The first and last rows of the image are the input image and Ground Truth, respectively, for comparison purposes. The second row shows the Grad-CAM visualizations of each network structure, and the third row displays the models' prediction results, with white boxes highlighting difficult-to-segment cell locations. Figure 7f demonstrates that our proposed AttEUnet effectively integrates edge prior information, focusing the model's attention on cell edges and interiors, as seen in the Grad-CAM images. The predicted masks in the third row have smooth edges and are complete in shape, with no internal voids. Figure 7d visualizes the Attention U-Net\* network without the Fusion Gate GF, lacking the ability to regulate the integration ratio of edge prior information, as reflected in the Grad-CAM heatmap showing unclear cell location recognition. Attention U-Net\* lacks sufficient detail recognition capability at the junction of two cell nuclei edges, failing to fully reflect the touching position of two cells. However, the MedT model's cell prediction results in Figure 7e appear fragmented, with visible straight line breaks at the seams near the junctions of two image patches, indicating that the small medical dataset is insufficient to train the transformer's global attention mechanism effectively. Figure 7a–c visualize the U-Net, UNet++, and Attention U-Net networks, which, thanks to the convolution structure's local receptive field and translational invariance priors, can correctly segment the approximate location of cells, although the masks often have internal voids and poor edge detail.



**Figure 7.** Qualitative analysis results for difficult-to-segment samples in the test set, with red boxes highlighting areas where AttEUnet outperforms other baseline networks. Note: “Attention U-Net\*” denotes the Attention U-Net model variant without the Attention-Enhanced branch and fusion module.

Additional predictive results for AttEUnet and baseline networks are displayed in Figure 8, where it is evident that our proposed network significantly surpasses baseline networks in terms of the smoothness and completeness of cell edge segmentation, with virtually no instances of missed target cells—a detail that IoU and F1 scores do not capture. For example, in Figure 8, the first and fourth rows highlight areas of dense cell populations within red boxes, where AttEUnet demonstrates smoother edges and fewer instances of cell edges sticking together. In the third and fifth rows of Figure 8, baseline networks exhibit severe cases of missed detections, with cell edges appearing fragmented and exhibiting irregular, jagged boundaries, whereas AttEUnet almost universally maintains complete cell segmentation forms. Our proposed network, by precisely controlling the multi-level fusion

process of edge prior feature maps with the original backbone feature extraction network through the Fusion Gate unit, significantly improves model accuracy on small datasets, promoting more efficient and precise utilization of small sample sets, effectively addressing the challenge of limited data availability.



**Figure 8.** Qualitative analysis results for sample test images from MoNuSeg and PanNuke datasets. The red box highlights regions where AttEUNet performs better than the other methods in comparison, making better use of edge attention mechanism. Note: “Attention U-Net\*” denotes the Attention U-Net model variant without the Attention-Enhanced branch and fusion module.

#### 4. Conclusions

In this work, we designed AttEUNet, an edge attention enhanced cell nucleus segmentation network for small medical datasets, addressing the challenges conventional segmentation networks face, such as difficulty in learning features and coarse edge segmentation effects on small medical datasets. Building upon the existing Attention U-Net network, we added an edge attention enhancement branch and fusion gate unit to improve the model’s segmentation accuracy on small medical sample sets. This method was trained and validated on two public datasets, MoNuSeg and PanNuke. The results show that compared to baseline networks, the AttEUNet model not only effectively enhances segmentation precision over the original network but also surpasses all baseline networks in segmentation accuracy. The smaller the dataset, the more pronounced the advantages of AttEUNet. Specifically, on the MoNuSeg and PanNuke datasets, it achieved F1 scores of 0.859 and 0.888, and IoU scores of 0.888 and 0.794, respectively. On the MoNuSeg dataset, which contains only 30 samples, it led the baseline networks with an F1 score of 0.027 to 0.176 and an IoU score of 0.019 to 0.18, confirming the effectiveness of our method. Our approach reached a foreground IoU score of 0.5 within just 15 epochs on the MoNuSeg dataset, which is at least 10.6 times faster than the convergence speed of baseline networks. This study offers a new network architecture for precise segmentation tasks on existing small medical datasets, alleviating the difficulty of training models on small medical datasets. The edge attention enhancement branch proposed in this research is highly generalizable and flexible, allowing for the integration of various prior attention mechanisms into the backbone network based on different tasks. Additionally, it can serve as an independent module added to other classic network structures. We provide new insights for segmentation tasks on small medical datasets, offering significant academic and practical value. Future work will explore the effects of the edge prior fusion branch on other network architectures to

further enhance the model's generalization ability. Moreover, using Grad-CAM images to reveal the internal workings of deep learning is only the first step; our model does not yet translate deep learning outputs into interpretable clinical decisions, especially for tasks more complex than lesion detection. It should intelligently offer recommendations to clinicians and provide reasons. We will focus more on model interpretability in the future to gain broader acceptance of AI as a clinical decision support tool.

**Author Contributions:** Conceptualization, K.M., Q.S., Z.W. and M.L.; Methodology, Y.Z., K.M., Q.S., Z.W. and M.L.; Software, Y.Z., K.M., Q.S., Z.W. and M.L.; Validation, Y.Z., Q.S. and Z.W.; Formal analysis, K.M.; Investigation, M.L.; Writing—original draft, Y.Z. and Q.S.; Writing—review & editing, Y.Z.; Supervision, K.M. and M.L.; Funding acquisition, M.L. All authors have read and agreed to the published version of the manuscript.

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

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data are available upon request.

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

## References

- Shen, D.; Wu, G.; Suk, H.-I. Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* **2017**, *19*, 221–248. [[CrossRef](#)] [[PubMed](#)]
- Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; van Ginneken, B.; Sanchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [[CrossRef](#)] [[PubMed](#)]
- Hesamian, M.H.; Jia, W.; He, X.; Kennedy, P. Deep learning techniques for medical image segmentation: Achievements and challenges. *J. Digit. Imaging* **2019**, *32*, 582–596. [[CrossRef](#)] [[PubMed](#)]
- Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
- Iglovikov, V.; Shvets, A. Terausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation. *arXiv* **2018**, arXiv:1801.05746.
- 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](#)] [[PubMed](#)]
- Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [[CrossRef](#)] [[PubMed](#)]
- 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.
- Wang, R.S.; Lei, T.; Cui, R.X.; Zhang, B.T.; Meng, H.Y.; Nandi, A.K. Medical image segmentation using deep learning: A survey. *LET Image Process.* **2022**, *16*, 1243–1267. [[CrossRef](#)]
- Liu, X.B.; Song, L.P.; Liu, S.; Zhang, Y.D. A Review of Deep-Learning-Based Medical Image Segmentation Methods. *Sustainability* **2021**, *13*, 1224. [[CrossRef](#)]
- Siddique, N.; Paheding, S.; Elkin, C.P.; Devabhaktuni, V. U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications. *IEEE Access* **2021**, *9*, 82031–82057. [[CrossRef](#)]
- Xiao, H.; Li, L.; Liu, Q.; Zhu, X.; Zhang, Q. Transformers in medical image segmentation: A review. *Biomed. Signal Process. Control* **2023**, *84*, 104791. [[CrossRef](#)]
- Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, 5–9 October 2015; pp. 234–241.
- Zhou, Z.W.; Siddique, M.M.R.; Tajbakhsh, N.; Liang, J.M. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In Proceedings of the 4th International Workshop on Deep Learning in Medical Image Analysis (DLMIA)/8th International Workshop on Multimodal Learning for Clinical Decision Support (ML-CDS), Granada, Spain, 20 September 2018; pp. 3–11.
- Alom, M.Z.; Yakopcic, C.; Taha, T.M.; Asari, V.K. Nuclei Segmentation with Recurrent Residual Convolutional Neural Networks based U-Net (R2U-Net). In Proceedings of the IEEE National Aerospace and Electronics Conference (NAECON), Dayton, OH, USA, 23–26 July 2018; pp. 228–233.
- Oktay, O.; Schlemper, J.; Le Folgoc, L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Y Hammerla, N.; Kainz, B.; et al. Attention U-Net: Learning Where to Look for the Pancreas. *arXiv* **2018**, arXiv:1804.03999.

17. Valanarasu, J.M.J.; Oza, P.; Hacihaliloglu, I.; Patel, V.M. Medical Transformer: Gated Axial-Attention for Medical Image Segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), Electr Network, Strasbourg, France, 27 September–1 October 2021; pp. 36–46.
18. Brigato, L.; Iocchi, L. A close look at deep learning with small data. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 2490–2497.
19. Romero, M.; Interian, Y.; Solberg, T.; Valdes, G. Training deep learning models with small datasets. *arXiv* **2019**, arXiv:1912.06761.
20. 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](#)]
21. Gamper, J.; Alemi Koohbanani, N.; Benet, K.; Khuram, A.; Rajpoot, N. Pannuke: An open pan-cancer histology dataset for nuclei instance segmentation and classification. In *Proceedings of the Digital Pathology: 15th European Congress, ECDP 2019, Warwick, UK, 10–13 April 2019*; Proceedings 15; Springer International Publishing: Berlin/Heidelberg, Germany, 2019; pp. 11–19.

**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.