BioMedInformatics

MDPI

Article

Deep Segmentation Techniques for Breast Cancer Diagnosis

Storm Schutte and Jia Uddin *

AI and Big Data Department, Endicott College, Woosong University, Daejeon 34606, Republic of Korea
* Correspondence: jia.uddin@wsu.ac.kr

Abstract: Background: This research goes into in deep learning technologies within the realm of medical imaging, with a specific focus on the detection of anomalies in medical pathology, emphasizing breast cancer. It underscores the critical importance of segmentation techniques in identifying diseases and addresses the challenges of scarce labelled data in Whole Slide Images. Additionally, the paper provides a review, cataloguing 61 deep learning architectures identified during the study. Objectives: The aim of this study is to present and assess a novel quantitative approach utilizing specific deep learning architectures, namely the Feature Pyramid Net-work and the Linknet model, both of which integrate a ResNet34 layer encoder to enhance performance. The paper also seeks to examine the efficiency of a semi-supervised training regimen using a dual model architecture, consisting of 'Teacher' and 'Student' models, in addressing the issue of limited labelled datasets. Methods: Employing a semi-supervised training methodology, this research enables the 'Student' model to learn from the 'Teacher' model's outputs. The study methodically evaluates the models' stability, accuracy, and segmentation capabilities, employing metrics such as the Dice Coefficient and the Jaccard Index for comprehensive assessment. Results: The investigation reveals that the Linknet model exhibits good performance, achieving an accuracy rate of 94% in the detection of breast cancer tissues utilizing a 21-seed parameter for the initialization of model weights. It further excels in generating annotations for the 'Student' model, which then achieves a 91% accuracy with minimal computational demands. Conversely, the Feature Pyramid Network model demonstrates a slightly lower accuracy of 93% in the Teacher model but exhibits improved and more consistent results in the 'Student' model, reaching 95% accuracy with a 42-seed parameter. Conclusions: This study underscores the efficacy and potential of the Feature Pyra-mid Network and Linknet models in the domain of medical image analysis, particularly in the detection of breast cancer, and suggests their broader applicability in various medical segmentation tasks related to other pathology disorders. Furthermore, the research enhances the understanding of the pivotal role that deep learning technologies play in advancing diagnostic methods within the field of medical imaging.

Keywords: deep learning; segmentation; breast cancer; feature pyramid network; linknet; student and teacher architecture

Citation: Schutte, S.; Uddin, J. Deep Segmentation Techniques for Breast Cancer Diagnosis. *BioMedInformatics* **2024**, *4*, 921–945. https://doi.org/10.3390/biomedinformatics4020052

Academic Editors: Hans Binder and Alexandre G. De Brevern

Received: 21 December 2023
Revised: 15 March 2024
Accepted: 22 March 2024
Published: 1 April 2024

1. Introduction

As technology increases, humans are becoming more capable of finding and diagnosing anomalies found in the body and the material world of things, particularly in cancerous tissues like breast cancer. These progressions in technology have brought forth mathematical equations like the perception introduced by Frank Rosenblatt et al., 1958 [1]. Although this formula shows simplicity, it is the foundation for more complex deep learning models such as Multilayer Perceptrons (MLPs) models as well as Convolutional Neural Networks (CNNs) which can further be seen in the well-known paper ImageNet classification with deep CNNs [2], which will be discussed further in this paper. These models have been the starting block to the deep learning world of artificial intelligence, and have allowed for the intelligence on medical images containing pathology anomalies through CNN architectures and segmentation.

Humanity faces many medical problems, one of which is that very prominent 'breast cancer' is detected through various imaging techniques, whether it be from Magnetic Resonance Imaging (MRI), computed tomography (CT) scans, or others. Humans have reached a point where they can detect medical disorders more effectively than before. The main part of this research will introduce breast cancer as one of the diseases that can be detected via segmentation and deep learning models. However, to do this, an understanding of cancer as well as training models are needed, to use the intelligence to further improve disease detection in the medical field.

To distinguish between healthy bodily tissues and those prone to disease or other abnormalities, segmentation can be used. Segmentation involves taking tiles of images from high-definition scans, as well as using masks or, rather, labeled data that identify regions of disease, to train the model to recognize what constitutes a disease or irregular tissue. This training is conducted via a deep learning model that performs millions of calculations to classify data points within a complex network of neurons.

2. Literature Review

In this section, the focus is on deep learning models and their understanding, specifically CNN and its influence on artificial intelligence. The methods these neural networks use to learn, including backpropagation and gradient descent, are explored. Additionally, the applications of CNNs in imaging, such as analyzing Whole Slide Images and segmenting images, are discussed. Lastly, the complexities involved in interpreting data using these models are addressed. However, as this paper is on breast cancer, some information should be more explored on this issue.

2.1. Global Disparities in Breast Cancer

Breast cancer poses a significant threat to women's health globally. In 2020, it afflicted approximately 2.3 million women worldwide and claimed 685,000 lives. Incidence rates are highest in developed regions such as Australia/New Zealand, Western Europe, and North America (>80 per 100,000 females), while the lowest rates are seen in parts of Africa, Asia, and Central America (<40 per 100,000). Sadly, some of the highest mortality rates (>20 per 100,000) occur in Melanesia, Western Africa, and Micronesia/Polynesia. These stark disparities highlight the urgent need for continued research into breast cancer [3]. Early detection offers the best chance of survival, and this is where tools like ultrasound (US) imaging come into play. Ultrasound is a painless and widely used technique that plays a crucial role in the early identification of breast cancer [4]. Please refer to Table 1 for the impact of breast cancer on a global scale.

Table 1. Breast cancer incidence (new cases) and mortality (deaths) in 2020 by world region and Human Development Index level [3].

	New Cases			Deaths		
Country	**N**	**ASR**	**Cum. Risk**	**N**	**ASR**	**Cum. Risk**
Eastern Africa	45,709	33	3.6	24,047	17.9	2
Middle Africa	17,896	32.7	3.4	9500	18	1.9
Northern Africa	57,128	49.6	5.1	21,524	18.8	1.9
Southern Africa	16,526	50.4	5.4	5090	15.7	1.7
Western Africa	49,339	41.5	4.5	25,626	22.3	2.5
Caribbean	14,712	51	5.5	5874	18.9	2
Central America	38,916	39.5	4.2	10,429	10.4	1.2
South America	156,472	56.4	6.1	41,681	14	1.5
Northern America	281,591	89.4	9.7	48,407	12.5	1.4
Eastern Asia	551,636	43.3	4.6	141,421	9.8	1.1

Table 1. *Cont.*

	New Cases			Deaths		
Country	**N**	**ASR**	**Cum. Risk**	**N**	**ASR**	**Cum. Risk**
All but China	135,265	66.9	7	24,247	9.4	1
China	416,371	39.1	4.2	117,174	10	1.2
South-Eastern Asia	158,939	41.2	4.5	58,670	15	1.7
South-Central Asia	254,881	26.2	2.9	124,975	13.1	1.5
All but India	76,520	27.5	3.1	34,567	12.9	1.5
India	178,361	25.8	2.8	90,408	13.2	1.5
Western Asia	60,715	46.6	5	20,943	16	1.7
Central-Eastern Europe	158,708	57.1	6.3	51,488	15.3	1.8
Northern Europe	83,177	86.4	9.4	17,964	13.7	1.5
Southern Europe	120,185	79.6	8.5	28,607	13.3	1.4
Western Europe	169,016	90.7	9.7	43,706	15.6	1.7
Australia/New Zealand	23,277	95.5	10.4	3792	12.1	1.3
Melanesia	2215	50.5	5.4	1121	27.5	2.9
Micronesia/Polynesia	381	58.2	6	131	19.6	2.1
Low HDI	109,572	36.1	3.9	58,586	20.1	2.2
Medium HDI	307,658	27.8	3	147,427	13.6	1.5
High HDI	825,438	42.7	4.6	247,486	12.1	1.4
Very high HDI		75.7	8.2	231,093	13.4	1.5
World		47.8	5.2	684,996	13.6	1.5

Female population; ASR = age-standardized rate per 100,000; Cum. Risk = cumulative risk, ages 0–74 years; HDI = Human Development Index.

These statistics underscore the critical importance of breast cancer and the necessity of adopting any possible measures to combat this disease. Early detection can significantly improve survival rates, highlighting the potential of deep learning models as a vital solution in this fight.

2.2. Deep Learning Models: Convolutional Neurol Network (CNN)

Deep learning models have made strides in the field of intelligence. They are a type of machine learning model that focuses on representation learning and uses neural networks. Unlike classical/traditional machine learning techniques, deep learning models excel at detecting and understanding patterns within large datasets by leveraging multiple layers of nonlinear processing. This ability has revolutionized how computers analyze and process data leading to possibilities in AI research and applications, and from the medical field to 3D objects in the metaverse [5–7].

One key aspect of learning models is their architecture, which consists of layers such as input, hidden, and output layers. Each layer performs operations on the input data, gradually transforming it. The 'deep' part refers to the presence of layers that enable the model to learn increasingly intricate features at each stage. The depth and breadth of these layers greatly impact the model's capacity to learn and adapt [8].

To enhance understanding of deep learning models and their various architectures, this research examined 61 deep learning models briefly, and they have been listed to help with the research in cancer and other pathologies (Table A1). Among them, two models emerged, LinkNet as well as FPN. To further comprehend the workings of these deep learning models, an exploration of backpropagation and gradient descent is essential.

2.3. Deep Learning Models, Backpropagation, and Gradient Descent

The learning process in deep learning models hinges on backpropagation and gradient descent. Backpropagation adjusts network weights based on output errors by applying the chain rule to compute the gradient of the loss function with respect to each weight, effectively propagating errors backward through the network [9]. Gradient descent, an iterative optimization algorithm, then minimizes the loss function by adjusting these weights based on the computed gradients [10]. Together, these mechanisms enable continuous

improvement of the model's predictions by minimizing the error between the predicted and actual outcomes.

2.4. Algorithm and Equations

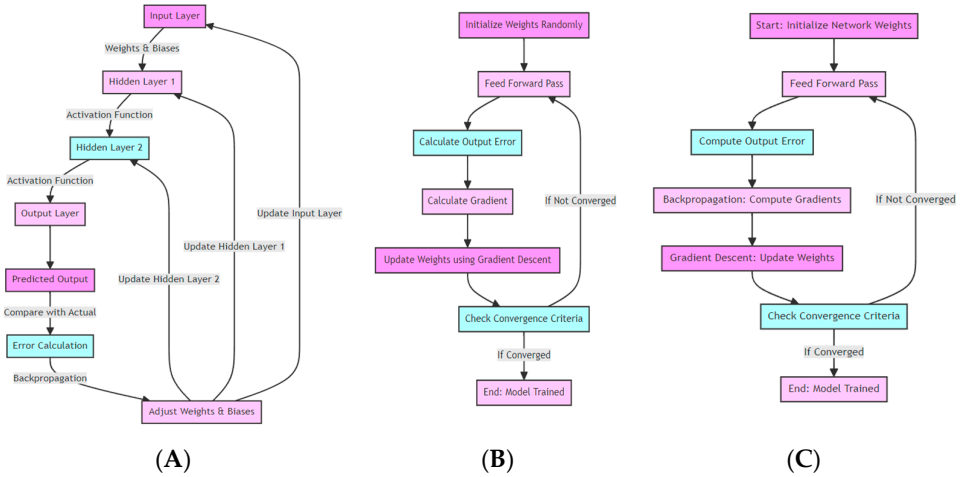
(I) Initialization: Start by selecting values, for the weights typically chosen randomly. (II) Compute Gradient: Employ backpropagation to calculate the gradient of the loss function concerning each weight. (III) Update Weights: Adjust the weights in the opposite direction to the gradient:

$$\text{Weight}_{\text{new}} = \text{Weight}_{\text{old}} - \eta \times \text{Gradient} \quad (1)$$

where η is referenced to the learning rate, as well as a small positive scalar determining the step size. (IV) Iterate: Repeat steps (II) and (III) until the loss function reaches its point indicating convergence [11].

2.5. Combining Backpropagation and Gradient Descent

In the context of neural networks, backpropagation and gradient descent (Figure 1C) work together. Backpropagation and gradient descent are two components in the field of neural networks. They collaborate to allow the model to learn from its mistakes effectively. Backpropagation calculates the gradients, providing the direction in which the weights need to be modified to minimize the loss. Subsequently, gradient descent utilizes these gradients to update the weights throughout iterations or epochs until the desired level of accuracy is reached [10].



(**A**) (**B**) (**C**)

Figure 1. (**A**) Illustrates Rumelhart 'backpropagation' [9]. (**B**) Illustrates gradient descent example [10]. (**C**) Illustrates backpfropagation and gradient descent [11].

2.6. Whole Slide Images (WSIs)

A lot of medical research is conducted using Visual Computing, and other similar types of research use WSIs. WSIs are high-resolution images that have been extracted from glass slides. WSIs become essential in the research used in pathology. WSIs are used in medical research, disease diagnosis, and education. One of the reasons for this is that WSIs allow for the entire tissue samples, which give pathologists the capability to examine and cross-reference the whole tissue, much like an improved version of microscopy [12,13].

WSIs are extensively used in CNN deep learning models. In this context, WSI has been used a lot in segmentation, classification, and detection. The reason for this is due to the size to which WSIs can extend, that being over gigabytes per image with a high level of specific details found in the image. This can bring challenges in computation and dealing with this large-size data effectively. Additionally, annotations of the data are time-consuming due to

their size and variability due to differences in staining slide preparation and the scanning across different labs [14–16].

An illustration of the WSI images is formed in a medical context. The image outlines the steps, from preparing the tissue sample to scanning and the creation of the digital image.

2.7. Segmentation

By definition, segmentation is defined as "the act of dividing something into different parts" [17]. Image segmentation, by definition, involves dividing images into segments or 'tiles' that are considered to be similarly related. Although the segmentation process depends on a low-level representation of data, which could relate to lighting, layer weight, and texture, segmentation cannot explicitly identify regions that may or may not belong to certain categories [18].

Semantic segmentation is distinct in that it classifies each pixel within an image, which is relevant because it allows for a more precise understanding of individual objects rather than just regions or segments. This method enables objects to be distinguished based on consistent patterns associated with them [18]. Semantic segmentation offers superior analysis compared to traditional segmentation methods. The patterns discerned in the images can provide insights into various problems encountered in the physical world.

2.8. Common Deep Learning Models Used for Cancer Detection

In the realm of breast cancer detection and diagnosis, the advent of deep learning models has marked a significant shift, particularly transforming medical imaging technologies. Notably, models such as U-Net, Mask R-CNN, and Generative Adversarial Networks (GANs) have emerged as landmarks, establishing new standards in the field. U-Net is celebrated for its precision in tumor detection and segmentation, demonstrating notable success in segmenting lung cancer lesions in PET/CT scans. Its convolutional encoding–decoding framework is particularly effective for intricate segmentation and improved classification accuracy, benefiting greatly from approaches like few-shot learning. This model's utility is not just confined to lung cancer but extends to breast cancer and various other diagnostic realms [19–21].

Following U-Net, Mask R-CNN has made substantial contributions by enhancing image analysis capabilities, especially in isolating and segmenting specific regions within medical imagery. This model's ability to perform pixel-perfect segmentation is particularly crucial for the complex visuals associated with breast cancer diagnostics, offering insights far beyond conventional methods. The precise demarcation of anatomical structures that Mask R-CNN facilitates is invaluable in the medical imaging sphere [22–24].

Generative Adversarial Networks (GANs) introduce a unique benefit by producing synthetic images to enlarge datasets, thereby addressing the dual challenges of data scarcity and privacy. This enhancement in data availability improves the training and performance of models without risking sensitive information. Additionally, GANs boost the robustness and precision of medical imaging models, significantly aiding the creation of more effective diagnostic tools [24,25].

This study further introduces the Feature Pyramid Network (FPN) and LinkNet models, selected for their specialized abilities to navigate the intricacies of different cancers detection. By examining these models in conjunction with the foundational progress made by U-Net, Mask R-CNN, and GANs, we undertake an analysis, shedding light on FPN and LinkNet.

Addressing the ongoing challenges, this research advocates for the integration of LinkNet and the Feature Pyramid Network (FPN). These models are distinguished by their effectiveness in managing the complex demands of medical image segmentation in different cancers detection. LinkNet, with its efficient segmentation process and lightweight architecture, promises to enhance the speed of tissue identification in high-resolution images without sacrificing accuracy, an attribute particularly beneficial for typical imaging datasets [26]. Conversely, FPN is recognized for its layered approach, significantly boosting

the model's capacity to detect breast cancer at various sizes, a feature crucial for identifying smaller, previously overlooked lesions, thereby advancing early detection efforts [27].

The selection of the LinkNet and Feature Pyramid Network (FPN) models designs is based on their ability to go beyond the limitations of current approaches. This allows for efficiency and flexibility. These models excel in handling variations in lesion sizes, which helps in creating thorough methods for detecting breast cancer. Moreover, this research aims to utilize a teaching strategy known as teacher–student learning to enhance the training process of these models.

In this teaching framework, a sophisticated 'teacher' model imparts knowledge to a simpler 'Student' model as discussed in this paper. The Student model then learns to imitate the performance of the Teacher model improving its capabilities with computational requirements. This method is especially useful when using complex models that are not feasible due to resource limitations. Through this approach, the Student model can efficiently and accurately detect breast cancer while maintaining effectiveness.

This study focuses on assessing how well the LinkNet and FPN models perform in terms of efficiency when using the teacher–student approach compared to established models. By exploring this teaching technique, we hope to discover strategies for enhancing detection methods and advancing breast cancer imaging further.

The combination of LinkNet and FPN, along with the teaching method using a teacher–student, approach shows potential for bringing insights and significant benefits which could accelerate advancements in identifying and treating breast cancer.

3. Methodology

In this section, the focus is on a training approach for segmentation models that handle data annotations. A dual-model architecture is employed, consisting of a 'Teacher' model (TM) and a 'Student' model (SM). The TM undergoes training on annotated WSIs using binary cross entropy and dice loss metrics to optimize performance. This model then generates soft pseudo annotations for WSIs to enhance the dataset. Subsequently, the SM is trained using both authentic and pseudo annotations, with a process that continues until maximum stability or convergence is achieved. Furthermore, the importance of evaluating the accuracy and overlap between the model's predictions and actual data is highlighted, using metrics such as the Dice Coefficient and Jaccard Index. These metrics are crucial in ensuring precision in medical imaging segmentation by aiming for similarity and complete overlap between predicted results and ground truth data.

3.1. Segmentation Model Training Strategy

Segmentation performed in machine learning, or without the means of segmentation, the semi-supervised training scheme offers an efficient strategy for images that have incomplete annotated data. The model used has a dual-model architect (Figure 2) which is the 'Teacher' model (TM) and a 'Student' model (SM). This process of training begins with the TM, where all the whole slide images (WSIs) in the dataset have annotated data or the equivalent masked data. This is then used to optimize the Binary Cross Entropy (BCE) as well as the Dice loss (DICE) as performance metrics [28].

The best performing model in validation is then selected as the best Teacher model (Figure 2). This model will then generate 'soft' pseudo annotations for the unlabeled WSI image data. It should be noted that although this method is not as trustworthy as true annotations, it enriches the dataset by a huge means [30].

The next step in this process is building on this expanded dataset. SM is then trained adding both authentic and pseudo annotations into its learning paths. The validation set again serves as the benchmark for selecting the better Student model, which, upon surpassing its teacher counterpart in performance, assumes the role of the teacher in subsequent iterations [31]. This iterative cycle is then cycled until the model's accuracy finds its highest stability, also known as a state of convergence. This iterative refinement takes advantage of the predictive accuracy of the TM to improve the SM, which overcomes

the problem of the lack of labeled data and increasingly improves the model's performance with each iteration [32].

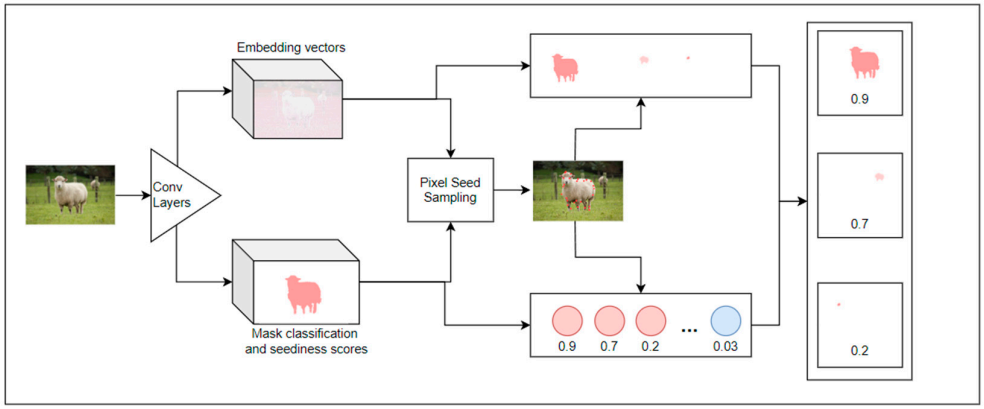


Figure 2. An image on the deep learning framework, from Fathi [29].

In this approach, the validation procedure plays a crucial role in ensuring the efficiency and dependability of segmentation models. After the training phase, each model—both TM and SM—undergoes a validation stage using a set of images not utilized during training. Throughout this stage, the models' performance is assessed on their capability to accurately segment images, which is vital for evaluating their ability to generalize. Performance is measured using the Dice Coefficient and Jaccard Index, which gauge the overlap between the models' predictions and the actual annotations. These validation metrics are essential as they offer an assessment of how the models' predicted segmentation aligns with ground truth data. A higher score signifies a close match between the models' output and the real data, which is critical in medical imaging where precision is key. The validation stage aids not only in choosing the best-performing models, but also in adjusting their parameters for optimal performance. By tracking these metrics, we can identify when a Student model has outperformed its teacher in terms of accuracy and generalization, guiding the training process towards generating more refined and accurate segmentation models. This methodical validation process guarantees that the ultimate model put into use is robust, trustworthy, and suitable for real-world use in diagnosis and treatment planning.

3.1.1. Dice Coefficient (Sørensen–Dice Index)

The Dice coefficient, which is also known as the Sørensen Dice index or Dice Similarity Coefficient (DSC), is a tool used in statistics. It is used to measure the commonness between two sets of data. In this research, this DSC is used to find the commonness of WSI segmentation. This is performed by DSC comparing the pixel-wise agreement between the ground truth segmentation and the segmentation shown by the deep learning model [33,34].

This is the formula used to prudence the dice coefficient:

$$DSC = |X| + |Y|2 \times |X \cap Y| \quad (2)$$

X is the predicted group of pixels that belong to the class of cancer; for example, Y, being the ground truth [34]. To iterate, an increased number in the DSC indicated an increased similarity between the prediction and ground truth. In this research, measurement is in the range of 0 to 1 [34].

3.1.2. Jaccard Index (Intersection over Union)

The Jaccard index, which is also known as the Intersection over Union (IoU), is, like the above, another tool used to measure the commonness in evaluating the performance of

the segmentation models. The formula is fairly simple, and is put together as the size of the intersection divided by the size of the union of the given sets:

$$JI = |X \cup Y||X \cap Y| \quad (3)$$

In this research, the Jaccard index has been used to quantify the percent overlap between WSIs given mask and the prediction output of these models. The range that is given in this research is again from 0 to 1, with 0 presenting no overlap and 1 presenting complete overlap. Complete overlap is the goal [35,36].

3.2. About the Dataset

The data used in this study, which is explained by the Kexin et al. [37] publication, consists of a collection of breast ultrasound images, formatted in .png (Figure 3) and the masked images in .mat format (Figure 4). These images are focused on breast cancer only and the images were approved by the institutional review board of Changhua Christian Hospital, in Taiwan. This retrospective and cross-sectional study includes 20,000 image tiles and 1,448,522 annotated nuclei. The total storage space required for this dataset is 6.8 GB. It is licensed under the CC BY license.

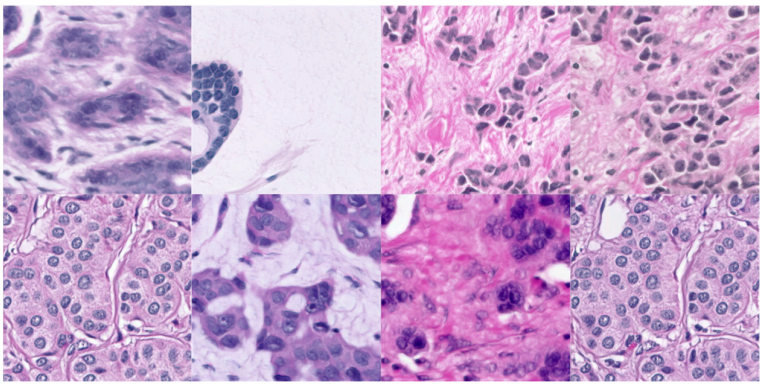


Figure 3. A view of the dataset is shown with cancerous tissue [37].

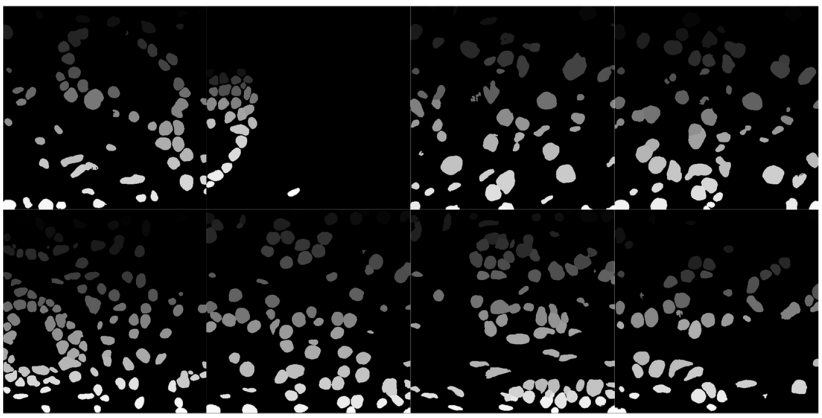


Figure 4. A view of the dataset labels [37].

The imaging process involved uses a GE Voluson 700 machine (manufactured by GE Healthcare based in Zipf, Austria). This machine ensured high-resolution captures with dimensions of 960 × 720 pixels in RGB mode (Figure 3). Each participant included in the study was between the ages of 35 and 75 and had been diagnosed with a malignant tumor through biopsy techniques. Each participant contributed two images taken from different scan plane angles [37].

In addition to the images themselves, this dataset also contains information about treatments including therapy methods using histology reports and radiography results. Furthermore, corresponding masks (Figure 4) are provided in. mat format.

To facilitate ease of access, for research purposes, and to enable studying breast tumor characteristics through imaging techniques effectively, all images and masks have been organized into a single folder. Furthermore, to enable this research to be more affective with the algorithms and code used in the research, the .mat files have been converted into .png files and reduced in size.

3.3. Deep Learning Models Being Compared

In this section, we will compare both deep learning models. To begin with, the layers of both models have been positioned side by side for an overall view. The model on the left is LinkNet (Figure 5A), and the model on right is FPN (Figure 5B).

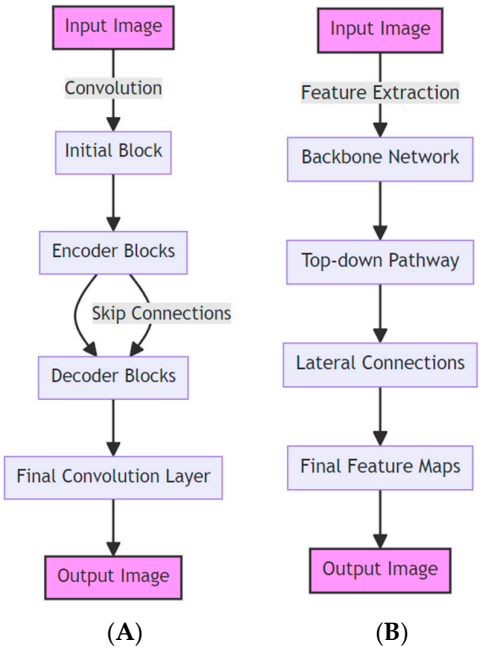


Figure 5. (**A**) An illustration of the LinkNet architecture [26]. (**B**) An illustration of the FPN architecture [38].

3.3.1. LinkNet

The introduction of the model known as LinkNet (Figure 5A) represents a significant advancement in the field of deep learning, specifically in efficiently segmenting images based on their semantic content. LinkNet is highly regarded for its ability to strike a balance between accuracy and computational requirements, making it an excellent choice for tasks that demand real-time performance. This efficiency can be attributed to its architecture, which utilizes trained encoders for extracting features and a decoder module for classification [26].

A noteworthy aspect of LinkNet is its design, which maintains a connection between the encoder and decoder modules. This connection plays a crucial role in mitigating the loss of resolution during down-sampling, enabling LinkNet to preserve information across the network with fewer parameters, leading to faster training without compromising performance [39].

LinkNet finds applications in various fields, such as medical imaging, where it aids in segmenting organs or lesions in radiographic images. It also excels at identifying road networks in aerial images. However, despite its versatility, LinkNet does have limitations. For instance, it may not perform as well as more complex models in handling fine details—especially critical in medical image segmentation where precision is paramount. Moreover,

the effectiveness of LinkNet heavily relies on the quality and diversity of the training dataset, which is a common limitation in machine learning models. Another challenge arises from the transferability of pre-trained models to medical images from non-medical domains, due to significant variations in image characteristics. Furthermore, although LinkNet is designed for real-time segmentation, it might encounter challenges when processing high-resolution imaging, which is typically found in this field [39].

3.3.2. Feature Pyramid Network

The Feature Pyramid Network (FPN) (Figure 5B) has emerged as a game-changing model in the field of image processing and analysis. Designed with sophistication, the FPN constructs a scale feature pyramid from a single input image. This complex design facilitates a top-down approach, where lateral connections enable the model to detect objects at various levels of the pyramid. This multifaceted detection process, as described by Lin [38], greatly enhances the model's ability to accurately segment and diagnose medical images.

FPNs have shown significant advancements in applications, particularly in object detection and segmentation. The pyramid structure of the network plays a crucial role in operating across scales, enabling the precise detection of fine details in medical images, as well as the identification of larger patterns. This capability was further explored in He's [22] study on Mask R-CNN, which utilized FPN for instance segmentation.

In terms of efficiency, FPNs offer an advantage over traditional methods used for multi-scale detection, as they process images much faster. The speed of real-time medical image analysis is pivotal, as discussed by Huang [40], who explore the trade-offs between speed and accuracy in convolutional object detectors. FPNs are incredibly valuable in time-sensitive settings due to their ability to quickly analyze complex images.

However, the FPN architecture does have its limitations. The added complexity and computational demands can be challenging in resource-constrained environments. Creating feature hierarchies, an aspect of FPN's design as explained by Kirillov [41], introduces inherent complexity. Furthermore, training an FPN can be difficult because it requires optimizing features across scales, a concept further explored by Guo [42] in their study on the SSD model, which serves as an alternative to FPN.

Another critical factor that impacts the effectiveness of FPNs is the quality of annotations used during training. The accuracy and reliability of medical image segmentation heavily rely on these annotations and are crucial for model deployment.

In summary, while FPNs bring groundbreaking advancements to object detection and segmentation in medical imaging, their complexity and dependence on high-quality annotations present challenges. It is crucial to grasp these subtleties to fully utilize the capabilities of FPNs in various applications. The studies mentioned above provide an understanding of FPNs' capabilities and the obstacles they face, offering valuable insights into the continuously advancing field of image processing technologies.

3.3.3. ResNet34 Layer Encoder Used in Both Models

The ResNet34 architecture is a variant of the Residual Network (ResNet) models, which are designed to enable the training of very deep neural networks. The '34' in ResNet34 denotes the use of 34 layers in the network, which includes convolutional layers, batch normalization layers, ReLU activations, and pooling layers. The encoder part of a ResNet34 model refers to the initial layers that are responsible for feature extraction from input images [43].

The ResNet34 encoder (Figure 6) has several limitations: (I) Computational Resources; the depth of the model necessitates computational resources, especially during training, which might not be feasible in all research or clinical environments. (II) Risk of overfitting; deeper models like ResNet34 can potentially overfit to the training data if proper regularization or data augmentation techniques are not implemented, particularly when dealing with less diverse datasets. (III) Challenges in optimization; while skip connections help

address the vanishing gradient problem, optimizing deep networks still poses difficulties. (IV) Adaptation to new tasks; although pre-trained ResNet34 encoders are readily available, customizing them for novel or highly specialized medical imaging tasks often requires fine-tuning efforts [19,40,43].

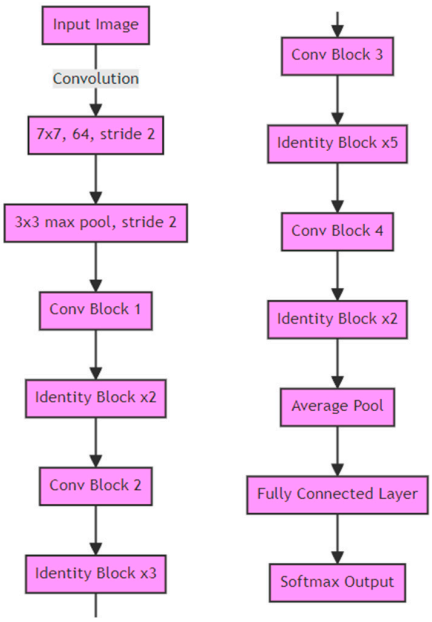


Figure 6. An illustration of the ResNet34 Layer Encoder [43].

3.4. Summary

In this section, the focus is on the research methodology that revolves around the training strategy for segmentation models, specifically handling images that have incomplete data. Our approach involves a semi-supervised training scheme using a dual-model architecture, consisting of a 'Teacher' model (TM) and a 'Student' model (SM). Initially, the TM is trained using WSIs that have annotated data, optimizing binary cross-entropy (BCE) and Dice loss (DICE) as performance metrics. The high-performing TM then generates soft pseudo annotations for the unlabeled WSI data, complementing the dataset even though these pseudo annotations are not as reliable as true annotations. Subsequently, the SM is trained using both authentic and pseudo annotations, with the validation set determining the better-performing Student model. This iterative process continues until our model reaches its highest stability or state of convergence, effectively addressing the challenge of limited labeled data while continuously enhancing our model's performance.

Moreover, in the research, two evaluation metrics are employed to assess the performance of our segmentation models: the Dice coefficient and the Jaccard index. The Dice coefficient measures the similarity between the predicted segmentation and the ground truth on a scale from 0 to 1, with a higher value indicating greater similarity. Similarly, the Jaccard index quantifies the extent of overlap between the provided mask and the model's output prediction. The goal is to achieve complete overlap for optimal performance. These statistical techniques are crucial in evaluating the precision and efficiency of the segmentation models when it comes to processing and analyzing WSIs.

4. Results Analysis

In this section, the focus is on meticulously comparing the performance of deep learning models such as the LinkNet and Feature Pyramid Network (FPN), all built on the ResNet34 framework. The analysis thoroughly assesses both Teacher and Student models under each architecture, evaluating their stability, accuracy, and image segmentation capabilities. This examination includes an exploration of seed values, training loss levels, validation Dice coefficients, and Jaccard indices. Through this evaluation, an understanding

of the strengths and limitations of these architectures in the field of medical image analysis is gained.

It should be highlighted that out of the 20,000 images, 5%, which corresponds to 1000 images, has been exclusively allocated for validation. Consequently, these 1000 images are excluded from the training dataset of the model. Despite the possibility of increasing this number, it was considered superfluous. This applies to both the LinkNet and FPN deep learning models. Additionally, both models were trained and validated on the same data, to keep a fair comparison.

4.1. Experiment with Linknet Architecture and Resent34 Base

4.1.1. Teacher Model Findings

The Teacher model of the LinkNet architecture using the ResNet34 base displayed different levels of training loss (Table 2, Figure 7) as well as validation metrics across three different seed values, 21, 42, 84. The research shows that the training loss ranges from 0.0098 to 0.0114, with an average of 0.9437. This would suggest that, despite the different seeds, the model's training was stable. Further looking into the validation Dice coefficient, the accuracy of the model in terms of perfectly overfitting the predicted image on the root image shows a 0.9425 and 0.9454 overlap rate, presenting strong accuracy.

Table 2. Linknet Architecture with ResNet34 Base: Teacher Findings.

Seed	Training Loss	Validation Dice	Validation Jaccard	Optimum Epoch	Figure Illustration
21	0.0102	0.9432	0.8930	92	N/A
42	0.0098	0.9425	0.8926	98	N/A
84	0.0114	0.9454	0.8972	85	Figure 7

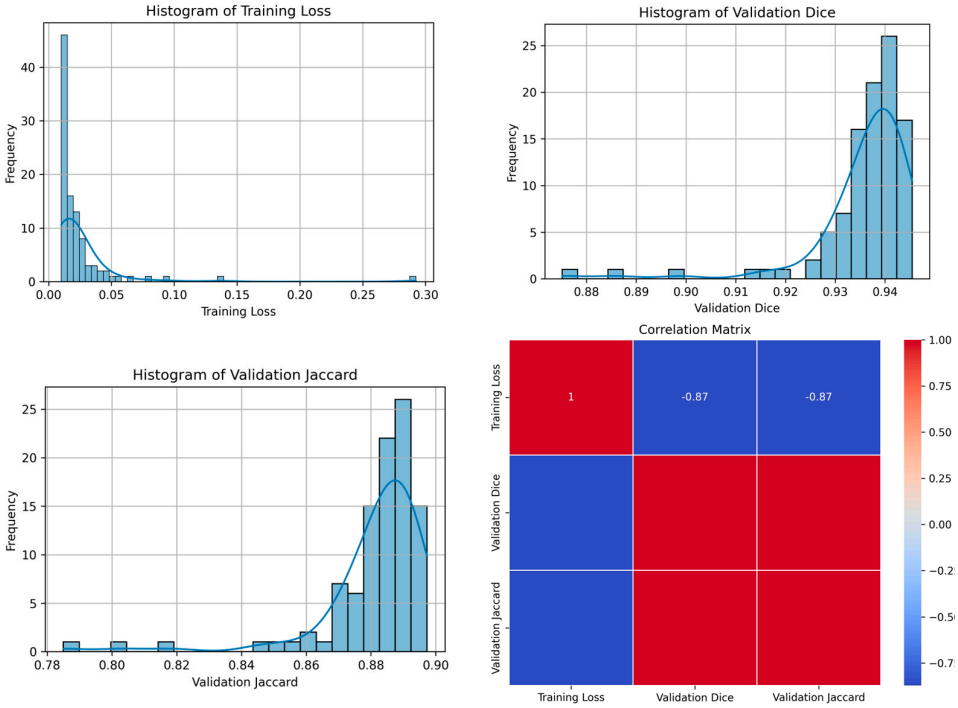


Figure 7. Results of the 'Teacher' model Linknet at seed 84.

However, further evaluating the Validation Jaccard index (Intersection over Union (IoU)), scores of 0.8926 to 0.8972 were produced. IoU presents an insight into the segmentation performance as well as indicating the accuracy of the predicted against the actual

segmentations. Although the scores are high, 89% would not necessarily be firm enough for real-world scenarios.

Further, the findings showed that the optimum epoch for the Teacher model is as follows. For a seed of 21, the model optimum epoch on 100 epochs was 92, seed 42 at 98 and finally 85 peaked at epoch 84.

4.1.2. Student Model Findings

The Student model is also based on LinkNet architecture with ResNet34, trained with the same seeds as the Teacher model. The Student model showed markedly lower loss than the Teacher model at 0.0004 to 0.0066 (Table 3, Figure 8); this suggests that the learning process was much more efficient or that there is potentially overfitting involved.

Table 3. Linknet Architecture with ResNet34 Base: Student Findings.

Seed	Training Loss	Validation Dice	Validation Jaccard	Optimum Epoch	Figure Illustration
21	0.0061	0.8938	0.8081	8	N/A
42	0.0066	0.8084	0.6784	9	N/A
84	0.0043	0.9529	0.9100	9	Figure 8

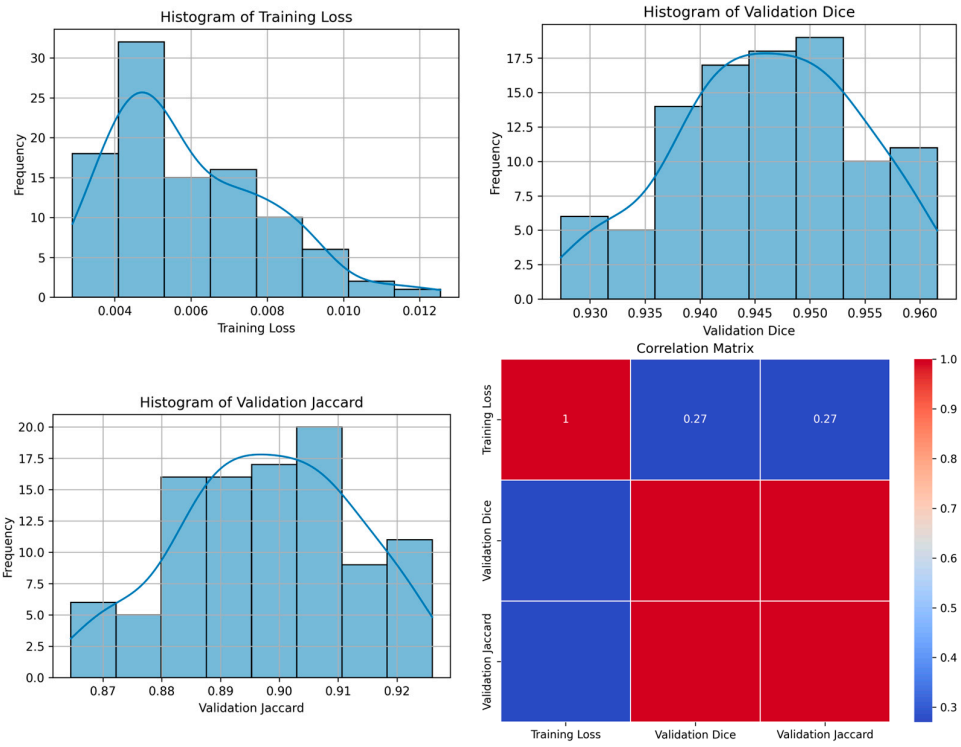


Figure 8. Results of the 'Student' model Linknet at seed 84.

The Validation Dice coefficient for the Student model shows a significant range from 0.8084 to 0.9529, with the highest score being achieved with seed 84. Similarly, the Validation Jaccard index varied from 0.6784 to 0.9100 with an average of 0.7988 as well as the highest score also associated with seed 84. This suggests that the initialization with seed 84 may have been particularly effective for the Student model. However, the Student model did not exhibit the same level of stability as the Teacher model, as evidenced by the fluctuation in the Validation Jaccard. This inconsistency could be attributed to initialization variance, wherein the initial weights may have converged to a less optimal region. This phenomenon is known to occur due to the initialization process or data shuffling, which are influenced by

the random seed assigned. These factors affect the learning process of the model, including the random sampling of data points or mini-batches.

Further looking into the optimum epoch for the Student model, the finding showed that the model reached much earlier than the teacher model, at epoch 9 for all seeds. This could imply that the Student model learned the necessary patterns more quickly than that of the Teacher model or that it may have begun to overfit after this point.

In summary, both the Teacher and Student models demonstrated strong performance on the given tasks (Table 4), with the Teacher model showing consistency across metrics and the Student model exhibiting rapid learning but with more variability in performance, displaying a lack of consistency in the Jaccard index. However, the use of different seed values also provided insights into the stability of the model's training process, with seed 84 yielding the best results for the Student model.

Table 4. Linknet Architecture with ResNet34 Base Comparison.

	Teacher			Student		
Seeds	**TL**	**VD**	**VJ**	**TL**	**VD**	**VJ**
21	0.0102	0.9430	0.893	0.0061	0.893	0.8081
42	0.0098	0.9425	0.8926	0.0066	0.8084	0.6784
84	0.0114	0.9450	0.8972	0.004	0.9529	0.9100
Average	0.0104	0.9435	0.8942	0.0055	0.8847	0.7988

Total Loss (TL), Validation Dice (VD), Validation Jaccard (VJ).

To better understand the results of the LinkNet architecture with ResNet34 encoder, Table 4 is presented comparing the Teacher and Student model results side by side. It is noticeable that the VD of the Teacher model is quite consistent, whereas the Student model shows some inconsistency, with seed 84 demonstrating the most promising accuracy.

4.2. Experiment with FPN Architecture and Resent34 Base

4.2.1. Teacher Model Findings

The Teacher model employing the Feature Pyramid Network (FPN) architecture with a ResNet34 base demonstrated varying training losses and validation metrics across three distinct seed values, 21, 42, and 84. The research indicated that the training loss fluctuated between 0.0173 and 0.0160 (Table 5, Figure 9), suggesting a relatively stable training process across different seeds. The validation Dice coefficient, which measures the accuracy of the model in terms of its overlap with the root image, ranged from 0.9326 to 0.9345. These rates indicate a strong accuracy in the model's predictions.

Examining the Validation Jaccard index, which offers insights into the segmentation performance, scores ranged from 0.8760 to 0.8795. Although these scores are high, they might not be sufficient for certain real-world applications. Regarding the optimum epoch, for seed 21, the best performance was achieved at epoch 78, for seed 42 at epoch 100, and for seed 84 at epoch 94.

Table 5. FPN Architecture with ResNet34 Base: Teacher Findings.

Seed	**Training Loss**	**Validation Dice**	**Validation Jaccard**	**Optimum Epoch**	**Figure Illustration**
21	0.0173	0.9331	0.8760	78	N/A
42	0.0154	0.9345	0.8780	100	Figure 9
84	0.0160	0.9326	0.8795	94	N/A

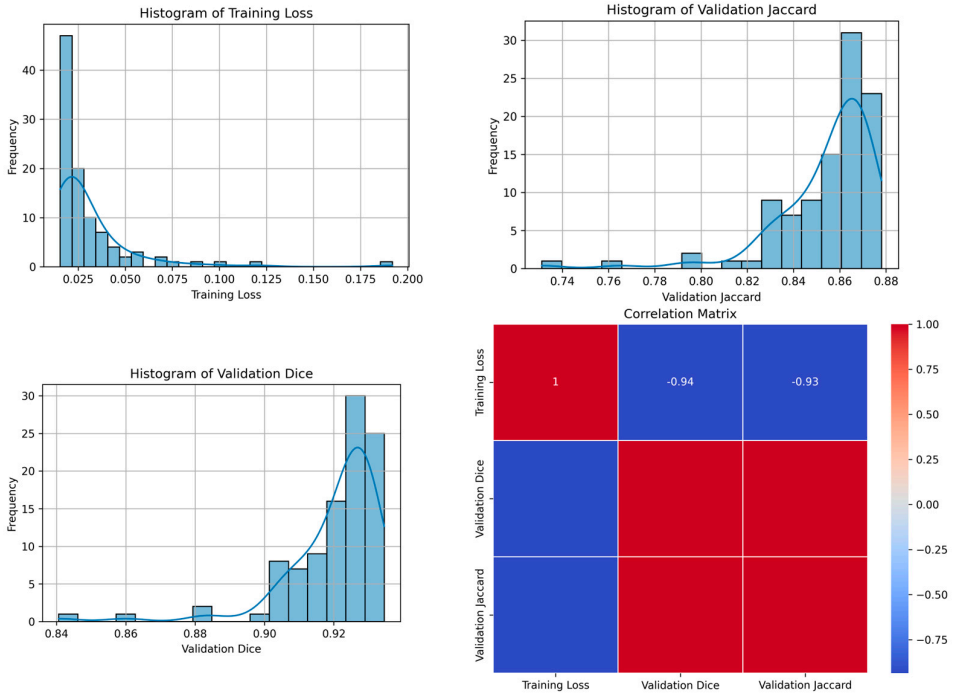


Figure 9. Results of the 'Teacher' model FPN at seed 42.

4.2.2. Student Model Findings

The Student model, also based on the FPN architecture with a ResNet34 encoder layer, trained with the same seed values as the Teacher model, showed varying degrees of training loss from 0.0183 to 0.0227 (Table 6, Figure 10). This indicates a variation in the learning efficiency, possibly hinting at overfitting in certain scenarios.

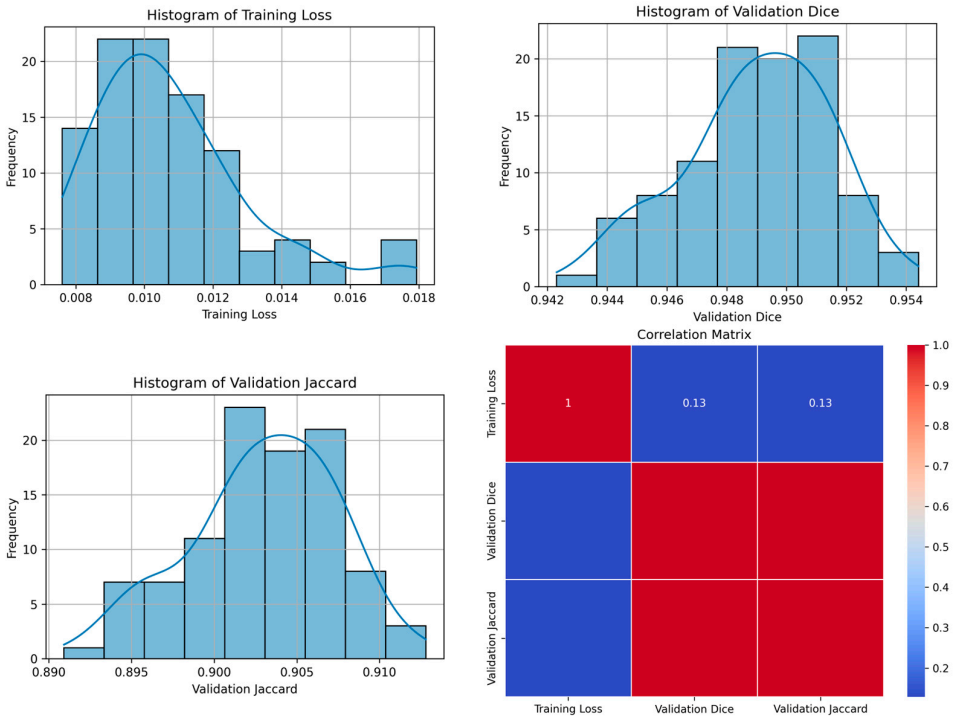


Figure 10. Results of the 'Student' model FPN at seed 42.

Table 6. FPN Architecture with ResNet34 Base: Student Findings.

Seed	Training Loss	Validation Dice	Validation Jaccard	Optimum Epoch	Figure Illustration
21	0.0183	0.8956	0.8109	2	N/A
42	0.0113	0.9544	0.9128	83	Figure 10
84	0.0227	0.9378	0.8830	2	N/A

The Validation Dice coefficient varied significantly, ranging from 0.8956 to 0.9544, with the highest accuracy achieved with seed 42. Similarly, the Validation Jaccard index ranged from 0.8109 to 0.9128, again showing a notable range in segmentation performance. The most effective seed for the Student model appeared to be 42, based on these metrics.

In terms of optimum epoch, the Student model reached its peak performance much earlier than the Teacher model, at epoch 2 for seeds 21 and 84, and epoch 83 for seed 42. This suggests a quicker learning curve for the Student model, although it may also indicate a tendency to overfit beyond these points.

In summary, both the Teacher and Student models under the FPN architecture with a ResNet34 base encoder layer demonstrated robust performance, with the Teacher model showing more consistency across metrics and the Student model exhibiting rapid learning, albeit with greater variability in its performance. The different seed values provided valuable insights into the stability and efficiency of the model's training process.

To better understand the results of the FPN architecture with ResNet34, Table 7 is presented, comparing the Teacher and Student model results side by side. It is noticeable that the VD of the Teacher model is consistent at 93%, whereas the Student model shows some inconsistency between 84 and 43 to 21, with seed 42 demonstrating the most promising accuracy.

Table 7. FPN Architecture with ResNet34 Base Comparison.

	Teacher			Student		
Seeds	**TL**	**VD**	**VJ**	**TL**	**VD**	**VJ**
21	0.0173	0.9331	0.0173	0.0183	0.8956	0.8109
42	0.0154	0.9345	0.0154	0.0113	0.9544	0.9128
84	0.0160	0.9326	0.0160	0.0227	0.9378	0.8830
Average	0.0162	0.9334	0.0162	0.0174	0.9292	0.8689

Total Loss (TL), Validation Dice (VD), Validation Jaccard (VJ).

4.3. Discussion

The primary objective of this study was to evaluate and compare the performance of two distinct deep learning architectures in analyzing Whole Slide Images (WSIs) with a focus on breast cancer segmentation. The results indicated that all models performed commendably overall. However, there were noticeable variations in terms of effectiveness and accuracy. Some models appeared to be overfitting the data, potentially limiting their generalizability. In contrast, Teacher models were generally more stable and accurate, while Student models learned quickly but exhibited more variation, resulting in a slightly lower quality compared to Teacher models.

The findings suggest that the LinkNet model shows strong performance in breast cancer detection, with its results being particularly notable with the seed 21 parameter. Although the FPN (Feature Pyramid Network) model did not match the performance of LinkNet, it still demonstrated sufficient utility in breast cancer detection as well as a more stable Jaccard index. This study highlights the potential of using different deep learning architectures for medical imaging tasks, with a focus on the effectiveness and efficiency of these models in detecting and analyzing breast cancer tissues.

It should be emphasized that these models have been developed using a teacher–student framework. In scenarios where ample data are available, the Teacher model

suffices, eliminating the necessity for a Student model. However, in cases of data scarcity, employing a Student model is recommended. According to the findings of this research, the Student model based on FPN performed satisfactorily. However, despite the varying outcomes where seeds 21 and 84 were tested, seed 42 exhibited inferior results, as previously discussed. This indicates that a Student model utilizing FPN can be effectively paired with the LinkNet teacher model. Nonetheless, it is important to consider that testing additional seed parameters could potentially modify the average performance metrics for a LinkNet-based Student model.

As seen in the images (Figures 11 and 12), the detection of breast cancer by each of the models was notably effective, demonstrating their capability to detect breast cancer in unseen data. This suggests that each model could be effectively utilized in detecting breast cancer and potentially in segmenting other pathological disorders in WSIs.

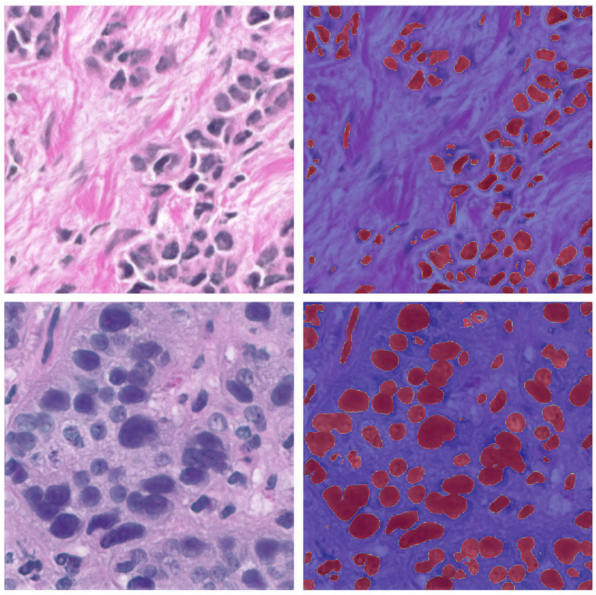


Figure 11. Image results detecting cancer with the Student model Linknet at seed 84.

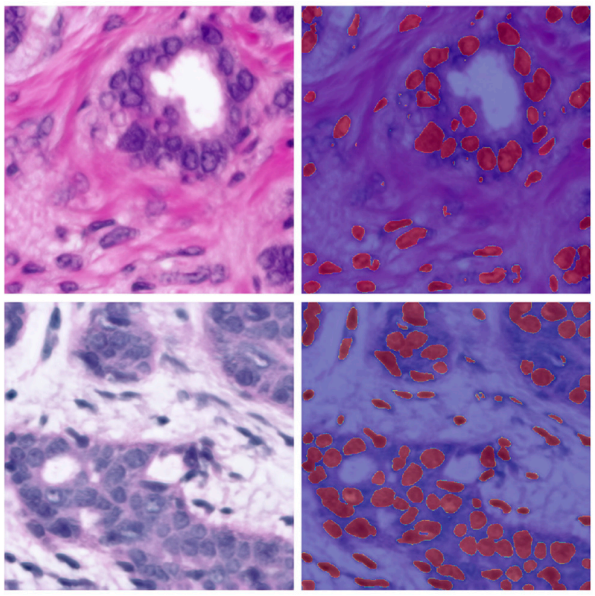


Figure 12. Image results detecting cancer with the Student model FPN at seed 42.

Finally, the use of Teacher and Student models demonstrated a strong approach to further train models on annotated data derived from the teacher model. This effectiveness was evident through the segmentation tools and the performance of the Jaccard and Dice evaluations.

4.4. Implications of the Study

In summary, this research emphasizes the significance of selecting and training models in deep learning tasks, setting important benchmarks for future advancements in image analysis technologies. Both individuals and organizations can greatly benefit from the increased accuracy and efficiency of these models, particularly in fields like medical imaging where precise image analysis is essential. These findings offer valuable guidance for organizations utilizing AI-based image analysis, aiding them in choosing and training models more effectively. On a broader scale, enhanced image analysis capabilities have significant implications for public safety, environmental monitoring, and policymaking, where accurate data interpretation is crucial. This study contributes to the existing body of literature on deep learning architectures in image segmentation by offering a perspective that can shape future research directions, particularly with a focus on optimizing model performance. Additionally, professionals working in AI and machine learning can utilize these insights to select and train models tailored to specific applications, thereby enhancing the practical usefulness of AI systems that rely on image analysis.

5. Conclusions

This research explores the performance of LinkNet and FPN architectures, all based on ResNet34, in image segmentation tasks for detecting breast cancer in Whole Slide Images (WSIs). The study provides critical insights into the stability of the training, accuracy levels, and segmentation capabilities of these models, making a significant contribution to the field of deep learning in medical image analysis.

The study highlights the Importance of selecting the appropriate model and seed value to optimize performance while addressing the challenges associated with overfitting, particularly in student models. These aspects are especially crucial in applications requiring precise image analysis, such as in the diagnosis of diseases like breast cancer or in environmental monitoring.

From a practical standpoint, these findings are invaluable in guiding model selection and refining training methodologies. Consequently, they have the potential to enhance the accuracy and efficiency of AI systems in healthcare, offering more effective tools for breast cancer detection and other medical imaging tasks.

Furthermore, this research demonstrates that deep learning architectures like LinkNet and FPN (Feature Pyramid Network) can accurately detect breast cancer in Whole Slide Imaging (WSI). Among these, LinkNet is highlighted as a recommended model due to its superior performance in this specific application. This investigation not only improves our understanding of deep learning models in image segmentation, but also underscores their potential to enhance diagnostic procedures in medical practice.

In summary, this study marks a significant advancement in the use of deep learning models for image analysis, particularly within the vital healthcare sector. It paves the way for more precise and reliable AI-powered diagnostic tools.

Author Contributions: Conceptualization, S.S.; methodology, S.S. and J.U.; software, S.S.; validation, S.S.; formal analysis, J.U.; investigation, S.S. and J.U.; resources, S.S.; data curation, S.S.; writing—original draft preparation, S.S.; writing—review and editing, S.S. and J.U.; visualization, S.S.; supervision, J.U.; project administration, J.U.; funding acquisition, S.S.; All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by Woosong University Academic Research 2024.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Datasets tried and used in this research are https://camelyon17.grand-challenge.org/ and https://zenodo.org/records/6633721.

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

Appendix A

This table presents a compilation of deep learning models reviewed during the course of this study. It is included to assist future researchers exploring deep learning models for image-based pathology detection.

Table A1. Table with 61 different deep learning model architectures.

No.	Year	Model (Architecture)	Description
1	2014	CRF Deep Learning	Combines a basic Convolutional Neural Network (CNN) with a Conditional Random Field (CRF) for improved image segmentation, enhancing boundary delineation by refining the CNN's output with the CRF [44].
2	2014	Fully Convolutional Networks (FCN)	Pioneering architecture in semantic segmentation that uses convolutional layers to process images of any size and outputs segmentation maps [45].
3	2015	U-Net	A highly effective network for medical image segmentation, featuring a U-shaped architecture that excels in tasks where data are limited [19].
4	2015	ReSeg	A model based on Recurrent Neural Networks (RNNs) and FCN, designed for semantic image segmentation, leveraging the sequential nature of RNNs for improved segmentation [46].
5	2015	Deconvolution Network	Uses deconvolutional layers to perform up-sampling of feature maps, enabling precise localization in semantic segmentation tasks [47].
6	2015	Dilated ConvNet	Incorporates dilated convolutions to expand the receptive field without reducing resolution, enhancing performance in dense prediction tasks like semantic segmentation [48].
7	2015	ParseNet	Enhances FCNs by adding global context to improve segmentation accuracy, focusing on understanding the whole scene context [49].
8	2015	SegNet	SegNet was designed for road scene understanding in the context of autonomous driving [50].
9	2016	DeepLab	DeepLabv1 and its successive versions (v2, v3, v3+, and v4) made significant contributions in semantic segmentation, incorporating dilated convolutions, atrous spatial pyramid pooling, and encoder–decoder structures [51].
10	2016	PSPNet	Proposed Pyramid Scene Parsing Network for scene parsing tasks [52].
11	2016	Instance-Aware Segmentaiton	This approach to segmentation is designed to not only classify pixels but also differentiate between separate instances of the same class in the image. It is commonly used in scenarios where identifying individual objects (instances) is crucial, such as in instance segmentation tasks [53].
12	2016	V-Net	An adaptation of the U-Net model for volumetric (3D) medical image segmentation. It is particularly effective for tasks like organ segmentation in 3D medical scans, using a similar architecture to U-Net but extended to three dimensions [54].
13	2016	ENet	A lightweight and efficient network designed for real-time semantic segmentation, particularly in mobile or low-power devices. It achieves a good balance between accuracy and speed, making it suitable for applications where computational resources are limited [55].
14	2016	RefineNet	Utilizes a multi-path refinement network for high-resolution semantic segmentation [56].

Table A1. *Cont.*

No.	Year	Model (Architecture)	Description
15	2017	Tiramisu	This is also known as The One Hundred Layers Tiramisu; it utilizes DenseNets for semantic segmentation [57].
16	2017	Mask R-CNN	An extension of Faster R-CNN, Mask R-CNN is effective for instance segmentation tasks [22].
17	2017	FC-DenseNet	Combines the principles of DenseNets (densely connected convolutional networks) with FCNs, leading to efficient and accurate semantic segmentation, especially in medical imaging [57].
18	2017	Global Convolutional Net	Designed for semantic segmentation, this network uses large kernels and global convolutional layers to handle both classification and localization tasks effectively [58].
19		DeepLab V3	An advanced version of DeepLab, it employs atrous convolutions and spatial pyramid pooling to effectively segment objects at multiple scales [59].
20	2017	FPN—Feature Pyramid Network	A versatile architecture used in both object detection and segmentation, it builds a multi-scale feature pyramid from a single input image for efficient and accurate detection at multiple scales [38].
21	2017	LinkNet	Utilizes an encoder–decoder architecture for fast and accurate semantic segmentation [26].
22	2018	ICNet	Designed for real-time semantic segmentation tasks [60].
23	2018	Attention U-Net	Incorporates attention mechanisms into the U-Net architecture [61].
24	2018	Nested U-Net	A U-Net architecture with nested and dense skip pathways [60].
25	2018	Panoptic Segmentation	A unified model that simultaneously performs semantic segmentation [62].
26	2018	Mask-Lab	A deep learning model that combines semantic segmentation, direction prediction, and instance center prediction for instance segmentation tasks [63].
27	2018	Path Aggregation Network	Enhances feature hierarchy and representation capability for object detection by enabling efficient multi-scale feature aggregation [64].
28	2018	Dense-ASSP	A network that integrates dense connections and atrous spatial pyramid pooling for robust semantic image segmentation [65].
29	2018	Excuse	A model that fuses semantic and boundary information at multiple levels to enhance feature representation and segmentation accuracy [63].
30	2018	Context Encoding Network	Focuses on capturing global context information for semantic segmentation, often using a context encoding module to improve performance [66].
31	2019	Panoptic FPN	A framework that combines the Feature Pyramid Network (FPN) with panoptic segmentation, effectively handling both object detection and segmentation tasks [41].
32	2019	Gated-SCNN	A semantic segmentation network with a gated shape stream that focuses on capturing shape information alongside the usual texture features [67].
33	2019	UPS-Net	A unified panoptic segmentation network that effectively combines instance and semantic segmentation tasks into a single, coherent framework [68].
34	2019	TensorMask	A dense prediction model for instance segmentation that uses structured 4D tensors to represent masks, enabling precise spatial understanding [69].
35	2019	HRNet	Maintains high-resolution representations through the network, enhancing performance in tasks like semantic segmentation and object detection [70].
36	2019	CC-Net: CrissCross Attention	Employs criss-cross attention to capture long-range contextual information in a computationally efficient manner for semantic segmentation [71].
37	2017	Dual Attention Network	Integrates position and channel attention mechanisms to capture rich contextual dependencies for improved scene segmentation [72].
38	2019	Fast-SCNN	A fast and efficient network design for semantic segmentation on road scenes [73].

Table A1. *Cont.*

No.	Year	Model (Architecture)	Description
39	2020	DPT	Vision transformer-based architecture for segmentation tasks [74].
40	2020	SETR	Another Vision Transformer-based method for segmentation shows the effectiveness of transformers in dense prediction tasks [75].
41	2020	PointRend	Aims at rendering fine-grained detail in segmentation through iterative subdivision [74].
42	2020	EfficientPS	Combines semantic segmentation and object detection efficiently [76].
43	2019	FasterSeg	An architecture search-based approach for real-time semantic segmentation [77].
44	2018	MAnet	Utilizes multi-head attention mechanisms for semantic segmentation [60].
45	2020	FasterSeg	FasterSeg is an AI-designed segmentation network that outperforms traditional models in speed and accuracy by using advanced neural architecture search and collaborative frameworks [78].
46	2020	PolarMask,	A novel single-shot instance segmentation method that represents object masks in a polar co-ordinate system; simplifies the instance segmentation process [79].
47	2020	CenterMask	An efficient anchor-free instance segmentation model that extends the CenterNet object detector by adding a spatial attention-guided mask branch [80].
48	2020	SC-NAS	Stands for "Semantic-Context Neural Architecture Search". It is a network architecture search method designed to optimize semantic segmentation networks by considering the semantic context of the task [81].
49	2020	EffientNet + NAS-FPN	This combines EfficientNet, a scalable and efficient network architecture, with NAS-FPN (Neural Architecture Search Feature Pyramid Network), a method for automatically designing feature pyramid architectures for object detection tasks. This combination aims to optimize both efficiency and accuracy in detection models [82].
50	2020	Multi-scale Adaptive Feature Fusion Network	Multi-scale Adaptive Feature Fusion Network for Semantic Segmentation in Remote Sensing Images [83].
51	2021	TUNet	TransUNet [84].
52	2021	SUnet	Swin-Unet, Swin-Transformer [85].
53	2021	Segm	Segmenter [86].
54	2021	MedT	Medical Transformer [87].
55	2021	BEiT	BERT Image Transformers [88].
56	2023	CrossFormer	A Hybrid Transformer Architecture for Semantic Segmentation [89]
57	2022	MLP-Mixer	Semantic Segmentation with Transformer and MLP-Mixer [90].
58	2022	Transformer-Powered Semantic Segmentation	Transformer-Powered Semantic Segmentation with Large-Scale Instance Discrimination [91].
59	2023	Adaptive Context Fusion	Semantic Segmentation with Adaptive Context Fusion [89].
60	2023	Multi-Scale Vision Transformers	Semantic Segmentation with Multi-Scale Vision Transformers [92]
61	2023	Hiformer: Hierarchical multi-scale	Semantic Segmentation with Hierarchical Vision Transformers [93]

References

1. Rosenblatt, F. The Perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **1958**, *65*, 386–408. [CrossRef] [PubMed]
2. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
3. Arnold, M.; Morgan, E.; Rumgay, H.; Mafra, A.; Singh, D.; Laversanne, M.; Vignat, J.; Gralow, J.R.; Cardoso, F.; Siesling, S.; et al. Current and future burden of breast cancer: Global statistics for 2020 and 2040. *Breast* **2022**, *66*, 15–23. [CrossRef] [PubMed]

4. Pacal, İ. Deep learning approaches for classification of breast cancer in ultrasound (US) images. *J. Inst. Sci. Technol.* **2022**, *12*, 1917–1927. [CrossRef]
5. Bengio, Y.; Courville, A.; Vincent, P. Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1798–1828. [CrossRef] [PubMed]
6. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [CrossRef]
7. Ananthachari, P.; Schutte, S. Big Data Tools, Deep Learning & 3D Objects in the Metaverse. In *Digitalization and Management Innovation II: Proceedings of DMI 2023*; IOS Press: Amsterdam, The Netherlands, 2023; Volume 376, p. 236.
8. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **2021**, *8*, 1–74. [CrossRef]
9. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [CrossRef]
10. Cui, N. Applying gradient descent in convolutional neural networks. In Proceedings of the 2nd International Conference on Machine Vision and Information Technology (CMVIT 2018), Hong Kong, China, 23–25 February 2018; Volume 1004, p. 012027.
11. Ruder, S. An overview of gradient descent optimization algorithms. *arXiv* **2016**, arXiv:1609.04747.
12. Kumar, N.; Gupta, R.; Gupta, S. Whole slide imaging (WSI) in pathology: Current perspectives and future directions. *J. Digit. Imaging* **2020**, *33*, 1034–1040. [CrossRef]
13. Pantanowitz, L.; Farahani, N.; Parwani, A. Whole slide imaging in pathology: Advantages, limitations, and emerging perspectives. *Pathol. Lab. Med. Int.* **2015**, *7*, 22–23. [CrossRef]
14. Campanella, G.; Hanna, M.G.; Geneslaw, L.; Miraflor, A.; Werneck Krauss Silva, V.; Busam, K.J.; Brogi, E.; Reuter, V.E.; Klimstra, D.S.; Fuchs, T.J. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat. Med.* **2019**, *25*, 1301–1309. [CrossRef] [PubMed]
15. Janowczyk, A.; Madabhushi, A. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *J. Pathol. Inform.* **2016**, *7*, 29. [CrossRef] [PubMed]
16. Litjens, G.; Sánchez, C.I.; Timofeeva, N.; Hermsen, M.; Nagtegaal, I.; Kovacs, I.; Hulsbergen-van de kaa, C.; Bult, P.; van Ginneken, B.; van der Laak, J. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Sci. Rep.* **2016**, *6*, 26286. [CrossRef] [PubMed]
17. Oxford Learner's Dictionaries. Segmentation Noun—Definition, Pictures, Pronunciation and Usage Notes. Available online: https://www.oxfordlearnersdictionaries.com/us/definition/english/segmentation (accessed on 20 December 2023).
18. Chowdhary, C.L.; Acharjya, D.P. Segmentation and feature extraction in medical imaging: A systematic review. *Procedia Comput. Sci.* **2020**, *167*, 26–36. [CrossRef]
19. 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; pp. 234–241.
20. Protonotarios, N.E.; Katsamenis, I.; Sykiotis, S.; Dikaios, N.; Kastis, G.A.; Chatziioannou, S.N.; Metaxas, M.; Doulamis, N.; Doulamis, A. A few-shot U-Net deep learning model for lung cancer lesion segmentation via PET/CT imaging. *Biomed. Phys. Eng. Express* **2022**, *8*, 025019. [CrossRef] [PubMed]
21. Falk, T.; Mai, D.; Bensch, R.; Çiçek, Ö.; Abdulkadir, A.; Marrakchi, Y.; Böhm, A.; Deubner, J.; Jäckel, Z.; Seiwald, K.; et al. U-Net: Deep Learning for Cell Counting, Detection, and Morphometry. *Nat. Methods* **2019**, *16*, 67–70. [CrossRef] [PubMed]
22. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
23. Soleimani, P.; Farezi, N. Utilizing deep learning via the 3D U-net neural network for the delineation of brain stroke lesions in MRI image. *Sci. Rep.* **2023**, *13*, 19808. [CrossRef] [PubMed]
24. Srinivasan, S.; Durairaju, K.; Deeba, K.; Mathivanan, S.K.; Karthikeyan, P.; Shah, M.A. Multimodal Bi-omedical Image Segmentation using Multi-Dimensional U-Convolutional Neural Network. *BMC Med. Imaging* **2024**, *24*, 38. [CrossRef]
25. Yi, X.; Walia, E.; Babyn, P. Generative adversarial network in medical imaging: A review. *Med. Image Anal.* **2019**, *58*, 101552. [CrossRef]
26. Chaurasia, A.; Culurciello, E. Linknet: Exploiting encoder representations for efficient semantic segmentation. In Proceedings of the 2017 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL, USA, 10–13 December 2017; pp. 1–4.
27. Zhang, H.; Xu, Z.; Yao, D.; Zhang, S.; Chen, J.; Lukasiewicz, T. Multi-Head Feature Pyramid Networks for Breast Mass Detection. In Proceedings of the ICASSP 2023—2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 4–10 June 2023; pp. 1–5.
28. Wang, Y.; Ahsan, U.; Li, H.; Hagen, M. A comprehensive review of modern object segmentation approaches. *Found. Trends Comput. Graph. Vis.* **2022**, *13*, 111–283. [CrossRef]
29. Fathi, A.; Wojna, Z.; Rathod, V.; Wang, P.; Oh Song, H.; Guadarrama, S.; Murphy, K.P. Semantic instance segmentation via deep metric learning. *arXiv* **2017**, arXiv:1703.10277.
30. Qiu, Z.; Gan, H.; Shi, M.; Huang, Z.; Yang, Z. Self-training with dual uncertainty for semi-supervised medical image segmentation. *arXiv* **2022**, arXiv:2304.04441v2.

31. Lin, Q.; Ng, H.T. A Semi-Supervised Learning Approach with Two Teachers to Improve Breakdown Identification in Dialogues. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2022; Volume 36, pp. 11011–11019. [CrossRef]
32. Sun, Z.; Fan, C.; Sun, X.; Meng, Y.; Wu, F.; Li, J. Neural semi-supervised learning for text classification under large-scale pretraining. *arXiv* **2020**, arXiv:2011.08626.
33. Bertels, J.; Eelbode, T.; Berman, M.; Vandermeulen, D.; Maes, F.; Bisschops, R.; Blaschko, M.B. Optimizing the DICE score and Jaccard index for medical image segmentation: Theory and practice. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2019, 22nd International Conference, Shenzhen, China, 13–17 October 2019. [CrossRef]
34. Zou, K.H.; Warfield, S.K.; Bharatha, A.; Tempany, C.M.C.; Kaus, M.R.; Haker, S.J.; Wells, W.M.; Jolesz, F.A.; Kikinis, R. Statistical Validation of Image Segmentation Quality Based on a Spatial Overlap Index. *Acad. Radiol.* **2004**, *11*, 178–189. [CrossRef] [PubMed]
35. Crum, W.R.; Camara, O.; Hill, D.L.G. Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE Trans. Med. Imaging* **2006**, *25*, 1451–1461. [CrossRef] [PubMed]
36. Taha, A.A.; Hanbury, A. Metrics for Evaluating 3D Medical Image Segmentation: Analysis, Selection, and Tool. *BMC Med. Imaging* **2015**, *15*, 29. [CrossRef]
37. Ding, K.; Zhou, M.; Wang, H.; Gevaert, O.; Metaxas, D.; Zhang, S. A Large-Scale Synthetic Pathological Dataset for Deep Learning-Enabled Segmentation of Breast Cancer. *Sci. Data* **2023**, *10*, 231. [CrossRef]
38. Lin, T.-Y.; Dollar, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 936–944. [CrossRef]
39. 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]
40. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1–9. [CrossRef]
41. Kirillov, A.; Girshick, R.; He, K.; Dollár, P. Panoptic Feature Pyramid Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 6399–6408.
42. Guo, C.; Fan, B.; Zhang, Q.; Xiang, S.; Pan, C. AugFPN: Improving Multi-Scale Feature Learning for Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 14–19 June 2020; pp. 12595–12604.
43. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1063–6919. [CrossRef]
44. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv* **2014**, arXiv:1412.7062.
45. Shelhamer, E.; Long, J.; Darrell, T. Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [CrossRef] [PubMed]
46. Visin, F.; Romero, A.; Cho, K.; Matteucci, M.; Ciccone, M.; Kastner, K.; Bengio, Y.; Courville, A. ReSeg: A recurrent neural network-based model for semantic segmentation. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 41–48. [CrossRef]
47. Noh, H.; Hong, S.; Han, B. Learning deconvolution network for semantic segmentation. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 13–16 December 2015. [CrossRef]
48. Yu, F.; Koltun, V. Multi-Scale Context Aggregation by Dilated Convolutions. *arXiv* **2016**, arXiv:1511.07122.
49. Wei, L.; Andrew, R.; Alexander, C.B. ParseNet: Looking Wider to See Better. *arXiv* **2015**, arXiv:1506.04579.
50. Qiao, J.-J.; Cheng, Z.-Q.; Wu, X.; Li, W.; Zhang, J. Real-time semantic segmentation with parallel multiple views feature augmentation. In Proceedings of the 30th ACM International Conference on Multimedia, Lisbon, Portugal, 10–14 October 2022; pp. 6300–6308. [CrossRef]
51. 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.* **2018**, *40*, 834–848. [CrossRef] [PubMed]
52. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890. [CrossRef]
53. Li, Y.; Qi, H.; Dai, J.; Ji, X.; Wei, Y. Fully Convolutional Instance-Aware Semantic Segmentation. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4438–4446. [CrossRef]
54. 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. [CrossRef]
55. Paszke, A.; Chaurasia, A.; Kim, S.; Culurciello, E. ENet: A Deep neural network architecture for real-time semantic segmentation. *arXiv* **2016**, arXiv:1606.02147. [CrossRef]

56. Lin, G.; Milan, A.; Shen, C.; Reid, I. RefineNet: Multi-path refinement networks for high-resolution semantic segmentation. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 5168–5177. [CrossRef]
57. Jegou, S.; Drozdzal, M.; Vazquez, D.; Romero, A.; Bengio, Y. The one hundred layers tiramisu: Fully convolutional DenseNets for Semantic segmentation. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017. [CrossRef]
58. Peng, C.; Zhang, X.; Yu, G.; Luo, G.; Sun, J. Large kernel matters—Improve semantic segmentation by global convolutional network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1743–1751. [CrossRef]
59. Chen, L.-C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic Image segmentation. *arXiv* **2017**, arXiv:1706.05587. [CrossRef]
60. 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.
61. Han, G.; Zhang, M.; Wu, W.; He, M.; Liu, K.; Qin, L.; Liu, X. Improved U-Net based insulator image segmentation method based on attention mechanism. *Energy Rep.* **2021**, *7*, 210–217. [CrossRef]
62. Xiong, Y.; Liao, R.; Zhao, H.; Hu, R.; Bai, M.; Yumer, E.; Urtasun, R. Upsnet: A unified panoptic segmentation network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 8818–8826.
63. Zhang, G.; Lu, X.; Tan, J.; Li, J.; Zhang, Z.; Li, Q.; Hu, X. Refinemask: Towards high-quality instance segmentation with fine-grained features. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 19–25 June 2021; pp. 6861–6869.
64. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8759–8768.
65. Eerapu, K.K.; Ashwath, B.; Lal, S.; Dell'Acqua, F.; Dhan, A.N. Dense refinement residual network for road extraction from aerial imagery data. *IEEE Access* **2019**, *7*, 151764–151782. [CrossRef]
66. Gu, Z.; Cheng, J.; Fu, H.; Zhou, K.; Hao, H.; Zhao, Y.; Zhang, T.; Gao, S.; Liu, J. Ce-net: Context encoder network for 2d medical image segmentation. *IEEE Trans. Med. Imaging* **2019**, *38*, 2281–2292. [CrossRef] [PubMed]
67. Takikawa, T.; Acuna, D.; Jampani, V.; Fidler, S. Gated-scnn: Gated shape cnns for semantic segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 5229–5238.
68. Cui, B.; Fei, D.; Shao, G.; Lu, Y.; Chu, J. Extracting raft aquaculture areas from remote sensing images via an improved U-net with a PSE structure. *Remote Sens.* **2019**, *11*, 2053. [CrossRef]
69. Chen, X.; Girshick, R.; He, K.; Dollár, P. Tensormask: A foundation for dense object segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 2061–2069.
70. Seong, S.; Choi, J. Semantic segmentation of urban buildings using a high-resolution network (HRNet) with channel and spatial attention gates. *Remote Sens.* **2021**, *13*, 3087. [CrossRef]
71. Huang, Z.; Wang, X.; Huang, L.; Huang, C.; Wei, Y.; Liu, W. Ccnet: Criss-cross attention for semantic segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 603–612.
72. Nam, H.; Ha, J.W.; Kim, J. Dual attention networks for multimodal reasoning and matching. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 299–307.
73. Poudel, R.P.; Liwicki, S.; Cipolla, R. Fast-scnn: Fast semantic segmentation network. *arXiv* **2019**, arXiv:1902.04502.
74. Chen, Z.; Zhu, Y.; Zhao, C.; Hu, G.; Zeng, W.; Wang, J.; Tang, M. Dpt: Deformable patch-based transformer for visual recognition. In Proceedings of the 29th ACM International Conference on Multimedia, Virtual, 20–24 October 2021; pp. 2899–2907.
75. Zheng, S.; Lu, J.; Zhao, H.; Zhu, X.; Luo, Z.; Wang, Y.; Fu, Y.; Feng, J.; Xiang, T.; Torr, P.H.S.; et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 19–25 June 2021; pp. 6881–6890.
76. Corradetti, M. A Point-Based Rendering Approach for On-Board Instance Segmentation of Non-Cooperative Resident Space Objects. Master's Thesis, Politecnico di Milano, Milan, Italy, 6 October 2022. Available online: https://hdl.handle.net/10589/195413 (accessed on 6 October 2022).
77. Mohan, R.; Valada, A. Efficientps: Efficient panoptic segmentation. *Int. J. Comput. Vis.* **2021**, *129*, 1551–1579. [CrossRef]
78. Chen, W.; Gong, X.; Liu, X.; Zhang, Q.; Li, Y.; Wang, Z. Fasterseg: Searching for faster real-time semantic segmentation. *arXiv* **2019**, arXiv:1912.10917.
79. Xie, E.; Sun, P.; Song, X.; Wang, W.; Liu, X.; Liang, D.; Shen, C.; Luo, P. Polarmask: Single shot instance segmentation with polar representation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 14–19 June 2020; pp. 12193–12202.
80. Lee, Y.; Park, J. Centermask: Real-time anchor-free instance segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 14–19 June 2020; pp. 13906–13915.
81. Song, Y.; Sha, E.H.-M.; Zhuge, Q.; Xu, R.; Xu, X.; Li, B.; Yang, L. Hardware-aware neural architecture search for stochastic computing-based neural networks on tiny devices. *J. Syst. Archit.* **2023**, *135*, 102810. [CrossRef]

82. Ghiasi, G.; Cui, Y.; Srinivas, A.; Qian, R.; Lin, T.Y.; Cubuk, E.D.; Le, Q.V.; Zoph, B. Simple copy-paste is a strong data augmentation method for instance segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 19–25 June 2021; pp. 2918–2928.
83. Shang, R.; Zhang, J.; Jiao, L.; Li, Y.; Marturi, N.; Stolkin, R. Multi-scale adaptive feature fusion network for semantic segmentation in remote sensing images. *Remote Sens.* **2020**, *12*, 872. [CrossRef]
84. Nguyen, V.A.; Nguyen, A.H.; Khong, A.W. Tunet: A block-online bandwidth extension model based on transformers and self-supervised pretraining. In Proceedings of the ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, 23–27 May 2022; pp. 161–165.
85. Fan, C.M.; Liu, T.J.; Liu, K.H. SUNet: Swin transformer UNet for image denoising. In Proceedings of the 2022 IEEE International Symposium on Circuits and Systems (ISCAS), Austin, TX, USA, 27 May–1 June 2022; pp. 2333–2337.
86. Strudel, R.; Garcia, R.; Laptev, I.; Schmid, C. Segmenter: Transformer for semantic segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Virtual, 19–25 June 2021; pp. 7262–7272.
87. Valanarasu, J.M.J.; Oza, P.; Hacihaliloglu, I.; Patel, V.M. Medical transformer: Gated axial-attention for medical image segmentation. In Proceedings of the Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, 27 September–1 October 2021; pp. 36–46.
88. Bao, H.; Dong, L.; Piao, S.; Wei, F. Beit: Bert pre-training of image transformers. *arXiv* **2021**, arXiv:2106.08254.
89. He, A.; Wang, K.; Li, T.; Du, C.; Xia, S.; Fu, H. H2Former: An efficient hierarchical hybrid transformer for medical image segmentation. *IEEE Trans. Med. Imaging* **2023**, *42*, 2763–2775. [CrossRef]
90. Lai, H.P.; Tran, T.T.; Pham, V.T. Axial attention mlp-mixer: A new architecture for image segmentation. In Proceedings of the 2022 IEEE Ninth International Conference on Communications and Electronics (ICCE), Nha Trang, Vietnam, 27–29 July 2022; pp. 381–386.
91. Razumovskaia, E.; Glavas, G.; Majewska, O.; Ponti, E.M.; Korhonen, A.; Vulic, I. Crossing the conversational chasm: A primer on natural language processing for multilingual task-oriented dialogue systems. *J. Artif. Intell. Res.* **2022**, *74*, 1351–1402. [CrossRef]
92. Mkindu, H.; Wu, L.; Zhao, Y. 3D multi-scale vision transformer for lung nodule detection in chest CT images. *Signal Image Video Process.* **2023**, *17*, 2473–2480. [CrossRef]
93. Heidari, M.; Kazerouni, A.; Soltany, M.; Azad, R.; Aghdam, E.K.; Cohen-Adad, J.; Merhof, D. Hiformer: Hierarchical multi-scale representations using transformers for medical image segmentation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–7 January 2023; pp. 6202–6212.