

Revolutionizing Disease Detection with U-Net: Enhancing Biomedical Image Segmentation in Healthcare

Zehao Fan *

Double-majored in Computer Science and Applied Statistics, The University of Toronto, Toronto, Canada

* Corresponding Author Email: zehao.fan@mail.utoronto.ca

Abstract. The growth of the global population and environmental pollution have led to an increased prevalence of diseases, putting significant pressure on medical systems worldwide. Manual detection methods are time-consuming and inefficient, highlighting the need for advanced solutions. U-Net, a robust deep-learning architecture designed for biomedical image segmentation, offers a transformative approach to human disease detection. Capable of generating its input data, the U-Net model requires reasonable training time and demonstrates high robustness, making it stable and suitable for real-world applications. This study explores the fundamental concepts and background of U-Net, emphasizing its practical application in patient care. Drawing from past studies, this paper underscores the urgency of integrating U-Net into healthcare for disease detection. It summarizes the impact and improvements of U-Net on biomedical image classification, identifies challenges in building and training the U-Net architecture, and proposes strategies for implementing this deep learning model in health systems to enhance disease detection.

Keywords: U-Net; Biomedical Image Segmentation; Deep Learning; Disease Detection.

1. Introduction

Most people will experience disease at some point in their lives, underscoring the imperative role of medical professionals in diagnosis and therapy. To ensure the highest accuracy in detection and the most effective therapy, many physicians rely on biomedical images to evaluate their patients' diseases and health conditions [1]. However, the manual reading and analysis of these images, such as computed tomography scans (CT), radiography, and ultrasound imaging, is a complex and time-consuming task that requires physicians to focus. There is a conflict between the high patient detection demand and slow manual detection speed.

As the population grows in cities, the demand for the health system to find diseases and cure them efficiently has become a challenge. Reading those biomedical images manually can be time-consuming. In addition, long periods of working with physicians may lead to an increase in the misdiagnosis rate. U-Net is a deep-learning convolutional neural network designed for biomedical-based segmentation and is renowned for its high efficiency and accuracy. It is an excellent tool for supporting physicians in detecting patients' diseases [2, 3].

The U-net is a dominant approach in biomedical image segmentation, and U-net has already achieved excellent results in biomedical segmentation [3, 4]. In addition, according to a study by Olaf et al., U-net is a brilliant architecture that performs excellently on various biomedical segmentation images within a reasonable training time (about 10 hours), and it only needs very few annotated images. It is generally believed that segmentation algorithms focusing on image processing, like the U-net, have been studied for a long time [5], so the U-net is mature enough to be used in the health system. Due to advancements in deep learning and machine learning, various computational techniques and architectures have been extensively used to segment biomedical photos [5]. In addition, computer vision models like U-Net architecture have become imperative in computer-aid diagnosis and image-guided clinical surgery because computer vision models usually have high accuracy and robustness. It is generally believed that Magnetic resonance imaging (MRI) is the best image to diagnose brain-related disease. However, since the MRI is complex and complicated to segment, to assist physicians in medical diagnosing, it is necessary to develop deep learning models

that can segment the soft tissue accurately and automatically [5]. Segmentation is one of the decisional components of quantitative analysis and disease diagnosing, but the traditional segmentation approach is expensive, complex, and inaccurate. Accordingly, it is impatient to introduce deep learning architectures like computer version models, which are known for being easy to train and highly accurate [5]. Segmentation techniques like U-Net in deep learning can find the clear tissue boundary in biomedical images. Also, segmentation can be used in tumour detection, cell counts, organ measurements, and mass detection. The UNet-Visual Geometry Group (VGG) 16 is one of the latest deep-learning models with outstanding performance on MRI image segmentation [5]. Specifically, UNet-VGG16 is proven accurate and quick to detect low-grade gliomas in the MRI of the fluid-attenuated inversion recovery sequence [5].

The primary purpose of this study is to convince people to use the U-net in the health system to support medical detection and deal with the challenges of a large number of medical detection requests because the U-net architecture is efficient and easy to train. The paper reviews and presents the foundational concepts and background of the U-net and compares manual detection and U-net image segmentation to show the advantages and high efficiency of the U-net. Also, Section II will detail U-net's features, structure, basic concepts and training steps. In Section III, this study will show the difficulties of unit the U-Net into the modern health system. Finally, section IV provides the conclusion.

2. Methodology

2.1. Dataset Description and Preprocessing

The Brats dataset is a widely used biomedical image dataset and a multimodal brain tumor image segmentation benchmark [6, 7]. The Brats dataset is a widely used biomedical image dataset and a benchmark for multimodal brain tumor image segmentation [7]. This dataset was utilized in the study conducted by Sahayam's group, which focused on brain tumor segmentation [6, 7]. For the Brats 2020, it contains 369 cases of brain cancer images. Specifically, 293 images are high-grade glioma, and 76 are grade glioma images. The dataset has four categories of images: the FLAIR, T1, T1ce and T2. The standard volume of those four categories are T1ce 1 mm, T2 weighted 2-6 mm, T1 (two layers) 1-6 mm and FLAIR 1-6 mm [6]. The Brats 2020 dataset generally uses the multi-institutional pre-operative MRI scan and focus. Also, the Brats 2020 dataset focuses on the image and data of intrinsically heterogeneous brain tumors, like gliomas [6].

2.2. Proposed Approach

U-Net is a convolutional neural network, a fundamental topic in deep learning. Its primary purpose is to maintain accuracy with less training data. This deep-learning convolutional neural network architecture consists of two main paths: contracting and expansive. The contracting path includes the encoder, which is responsible for learning contextual information, while the expansive path contains the decoder, which decodes the data output by the encoder. U-Net performs exceptionally well even with small input datasets, thanks to data augmentation techniques applied to the input data to generate new data. Standard augmentation techniques include rotating, flipping, and scaling. The U-Net model can use these techniques to create new training data from existing data [1, 2]. U-Net is a powerful segmentation model that processes input images and produces new images using data augmentation techniques. This capability allows U-Net to generate more input data, ensuring robustness in image segmentation tasks even with a limited dataset.

2.2.1. Introduction to U-Net

The contracting path captures critical information from the input data. Specifically, the encoder inside the contracting path usually utilizes convolutional techniques, which can capture more details of the input data. The contracting path performs similarly to the feedforward layers of other

convolutional neural networks. The expansive path's vital structure is that the decoder layers usually upsample the feature—maps using the convolutional operation.

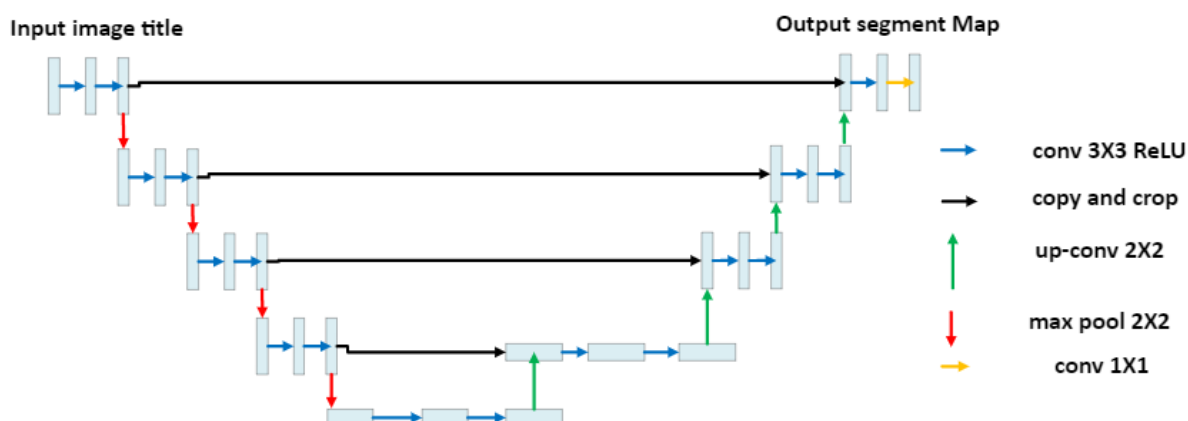


Fig 1. The Architecture of the U-Net.

Fig. 1 is the basic structure of a U-Net, and each blue rectangle inside the chart is a multi-channel feature map. There are four encoders on the architecture's left and four decoders on the model's right. Each encoder reviews the input image to extract useful information using the convolutional layers during the encoding process. In addition, during the decoding process, the encoders always upsample the feature using the transpose convolution. There are direct connections between some of the encoders and decoders. Specifically, those black arrows are called skip connections, responsible for concatenating the output of encoders and decoders.

Usually, the input to the encoder is the original images, and the decoder's output is the predicted mask. Skip connections between the encoders and decoders bypass some layers in the neural network and feed the output of one layer directly as the input to subsequent layers. The skip connections help to transfer the selected feature directly from the encoders to the decoders part of the model. Finally, the skip connection supports the decoder in generating a better segmentation mask. The connections between two encoders are called max pooling, and some layers are used to down-sample the input. After each max pooling, the channels are doubled. The decoder is a verse of the encoder. An up-sampling operation between each encoder can restore the special resolution of the features lost during the encoding phase. Usually, the encoder feature contains more special features, and the decoder feature might include more semantic information. The U-Net structure can produce pixel-perfect segmentation after merging the decoder and encoded feature.

2.2.2. Introduction to attention enhanced U-net with hybrid dilated convolution (AE U-Net with HDC)

The U-Net is highly influential in biomedical segmentation, particularly in detecting breast cancer [1]. Breast ultrasound images are typically input into the U-Net model [8]. However, different breast tumors' varying shapes and sizes in ultrasound images can reduce the model's accuracy and affect segmentation performance [8]. To address this issue, Yu Yan and colleagues developed an enhanced version of U-Net, called Attention Enhanced U-Net with Hybrid Dilated Convolution (AE U-Net with HDC), to improve the segmentation accuracy of tumors with diverse appearances. Unlike the standard U-Net, this improved model incorporates a new parameter called the dilation rate, which denotes the spacing between values in the convolutional kernel. Using dilated convolution ensures an expanded receptive field without reducing the resolution of the feature image. Compared to the ordinary U-Net, AE U-Net with HDC is more accurate in breast cancer-related image segmentation, effectively identifying and focusing on the tumor region. The dilated convolution improves receptivity without affecting feature image resolution, enhancing the U-Net model's performance in image segmentation tasks [1].

AE U-Net with HDC, like the standard architecture, contains three main parts: the contraction path, the expansion path, and three Attention Gates (AGs). In AE U-Net, each AG focuses on the lesion

part of the image and the background. Additionally, a new loss function is used to update the weights of the contraction path, AGs, and expansion path. This new loss function utilizes the lesion region mask to update the weights, increasing the U-Net's performance. The training method also changes, employing a partial loss function to train the leading network and AGs. In the ordinary U-Net, downsampling reduces the image size, and upsampling recovers it, which might lead to information loss. Several groups of HDCs replace standard convolution operations to address this, mitigating information loss and improving image segmentation production. Ultimately, the AE U-Net with HDC model enhances image segmentation accuracy and can effectively support medical detection in the healthcare system [8].

2.2.3. Introduction to ResNet-based U-Net

Brain cancers can harm patients' health and lives, so it is necessary to detect brain tumors and cancer quickly and accurately [9]. San's group trained a ResNet-based U-Net for brain tumor MRI image segmentation to aid in brain tumor detection [10].

This ResNet-based U-Net employs three preprocessing methods. Normalization of MRI slices, extraction of paths, and data augmentation [10]. It comprises seven residual units, 25 convolutional layers, and 15 batch normalization layers. The residual blocks inside the ResNet-based U-Net help deal with degradation problems. By skipping connections between layers, the model reduces the degradation problem. Compared to the ordinary model, the ResNet-based U-Net performed excellently in core tumor-related image segmentation. The ResNet-based U-Net (R U-Net) contains three parts: the contraction path on the left-hand side, the expanding path on the right-hand side and the bottleneck between them. Each encoder is concatenated by three parts: one residual block substitute into two plain blocks. Each decoder has the same structure as an encoder [5, 9, and 10]. The residual block contains a convolutional layer with a 2X2 filter and uses the ReLU activation function. In addition, the residual block in the bottleneck acts as the bridge between the expanding path and the expanding path. Each encoder downsamples the features, and each decoder upsamples them. The softmax activation function is also applied in the last layer to reduce dimensionality. In contrast to the ordinary U-Net, the upgraded model performs better in image segmenting and finding tumour areas.

3. Result and Discussion

3.1. Result Analysis

According to Table 1, the result of Yu Yan et al.'s study, the AE U-Net with HDC has the highest data in all seven-performance metrics. The data for AE U-Net are all higher than the data for U-Net. The AE U-Net with HDC has the best performance, and the AE U-Net ranks second, while the ordinary U-Net has the worst performance.

Table 1. The Result of U-Net, AE U-Net and AE U-Net with HDC.

	Precision	Recall	F1 Score	AUC	Acc	M-IOU	Dice
U-Net	83.34 ±	77.02 ±	80.06 ±	94.75 ±	93.55 ±	79.78 ±	88.75 ±
	0.75%	3.52%	1.74%	1.58%	1.34%	1.53%	1.16%
AE U-Net	86.28 ±	78.69 ±	81.96 ±	95.54 ±	94.18 ±	81.81±	89.99 ±
	0.61%	0.79%	1.47%	2.19%	0.32%	1.20%	0.84%
AE U-Net with HDC	87.53 ±	80.48 ±	83.03 ±	97.14 ±	95.81 ±	83.20 ±	90.83 ±
	0.79%	0.68%	1.19%	1.38%	1.39%	0.92%	0.58%

Since the AE U-Net's data in all performance metrics are higher than the data in the same metrics of the ordinary U-Net model, the result strongly proves that attention enhancement can raise the robustness of the U-Net model. Similarly, the data of AE U-Net with HDC (See in Fig. 2) are all higher than that of AE U-Net in the same category, so the hybrid dilated convolution layers lead to better performance of the AE U-Net compared to the regular convolution layers.

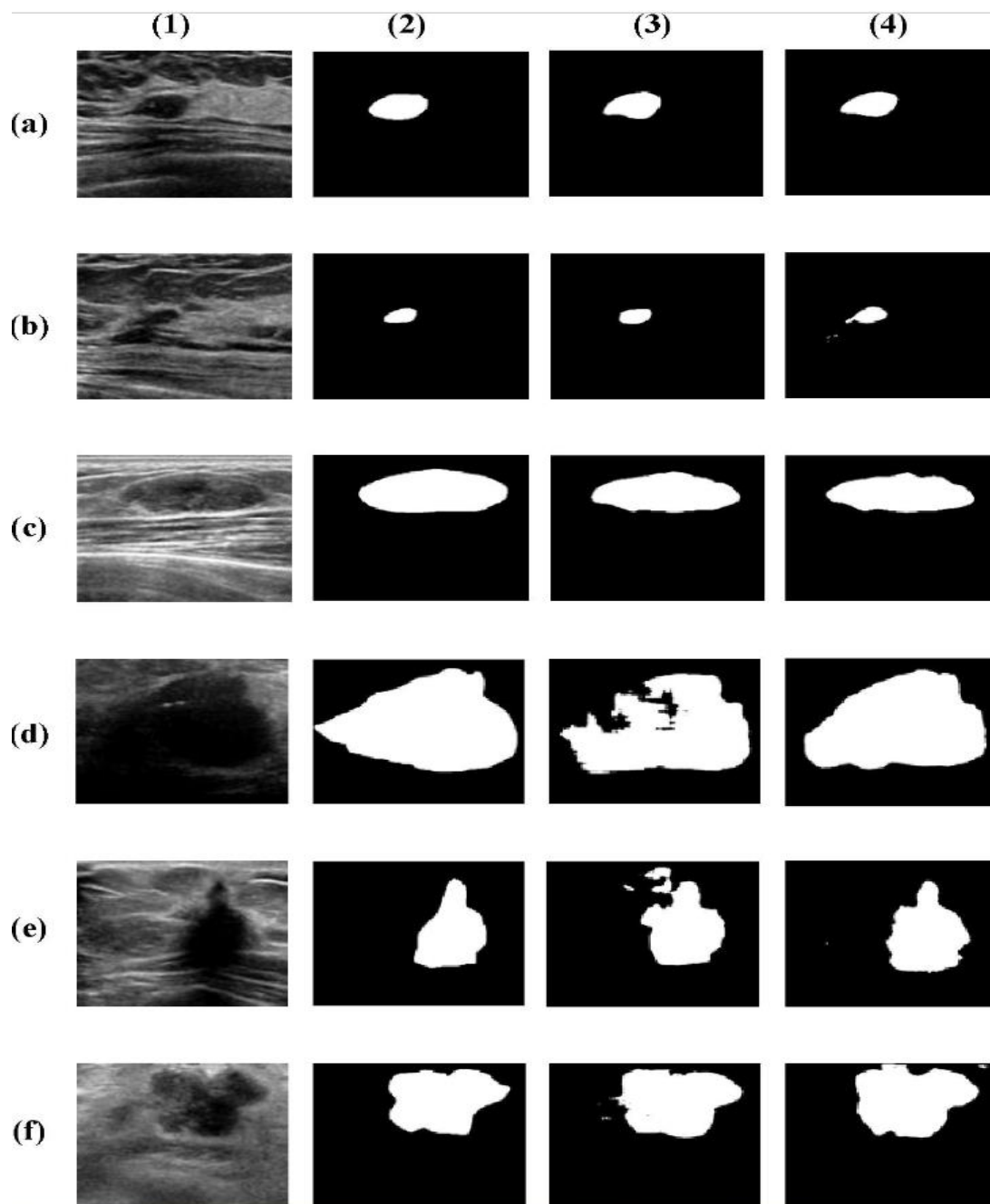


Fig 2. The Segmentation output of AE U-net and AE U-net with HDC.

The first column of those images is ultrasound breast images. The second column is the mask. The Third column is the segmentation output of the AE U-Net. The last column is the segmentation output of the AE U-net with HDC. Comparing the output of AE U-net and the output of AE U-net with HDC, the outputs of the model with HDC are more similar to the mask comparing the model without HDC. In addition, the results of the model with HDC retain more information and details than the output of the model without HDC. The images in rows (d) and (e) can be excellent examples of the model's output, with HDC being better quality than without HDC. Especially for row (d), the output of the AE U-Net has some noticeable loss in the data collection, so there are some dark shadows. However, there are no shadows in the output of the AE U-Net with HDC, which means outputs of AE U-Net with HDC have superior segmentation results.

3.2. The benefits and drawbacks of AE U-Net with HDC

According to Yu Yan et al.'s study outputs, AE U-Net with HDC is more accurate in image segmentation than ordinary U-Net, especially when the input images are ultrasound breast images.

Also, the AE U-Net with HDC retains the advantages of the ordinary U-Net. It is easy to train. Specifically, since U-Net can produce some training data by itself, producing high-quality image segmentation requires a reasonable amount of training data. In addition, it provides efficient and high-quality image segmentation on breast ultrasound photos.

Furthermore, the model's output keeps most of the information from the original images, which means it is an excellent tool in health systems. However, since AE U-Net with HDC is only trained on breast ultrasound images, it may not perform well in other biomedical images. In addition, like other U-Nets, the model requires all input images to have the same size to perform high-quality segmentation. However, in a real-world hospital scenario, the input biomedical images may have a different size, which might reduce the model's image segmentation performance.

3.3. Improvement and Expectations

AE U-Net with HDC is still a great tool after a few improvements. Since there are hundreds of biomedical image categories, outstanding performance only in ultrasound breast image segmentation is insufficient. To increase its utility, the model should be trained on more general categories of biomedical images. Also, to improve its robustness, the model should be improved to adapt to different sizes of input images so that this model can be widely used in various health systems. Although AE U-Net with HDC still has some disadvantages, it can help physicians with medical detections. Since it is easy to use and efficient, it can eliminate the conflict between limited manual detection speed and many patients. Also, the model is known for its high accuracy in image segmentation and excellent stability, which can reduce the probability of Medical Malpractice due to physician fatigue. Eventually, the AE U-Net with HDC will be applied worldwide due to its high efficiency, stability and robustness.

4. Conclusion

The study highlights the strain on healthcare systems worldwide due to the conflict between limited manual detection speed and the high demand for efficient patient diagnosis. To address this, new tools and techniques must be introduced to support physicians in improving detection speed. U-Net, a model primarily used in biomedical segmentation, is known for its high accuracy, stability, and reusability. Additionally, U-Net's ability to generate its training data makes it cost-effective. This study introduces two advanced versions of U-Net: the AE U-Net with HDC and the ResNet-based U-Net. The AE U-Net with HDC excels in breast cancer segmentation, effectively identifying and focusing on tumor regions due to its use of dilated convolution, which enhances segmentation performance.

On the other hand, the ResNet-based U-Net, developed by San's team, is optimized for brain tumor MRI image segmentation and detection. This model addresses the degradation issues of the ordinary U-Net by skipping some connections between layers, resulting in improved performance. In conclusion, U-Net is a mature deep-learning model suitable for integration into medical systems and biomedical image segmentation. Implementing U-Net in healthcare can alleviate the burden of limited manual detection speeds and the high volume of patients, ultimately reducing medical malpractice.

References

- [1] Yan Y. Liu Y. Wu Y. et al. Accurate segmentation of breast tumors using AE U-net with HDC model in ultrasound images. *Biomedical Signal Processing and Control*, 2022, 72: 103299.
- [2] ORonneberger P.F. TBrox: U-Net: Convolutional networks for biomedical image segmentation. *MICCAI*, 2015: 234-241.
- [3] Sahiner B. Chan H.P. Hadjiiski L.M. et al. Computer-aided detection of clustered microcalcifications in digital breast tomosynthesis: a 3D approach. *Medical physics*, 2012, 39(1): 28-39.

- [4] Cao H. Wang Y. Chen J. et al. Swin-unet: Unet-like pure transformer for medical image segmentation. European conference on computer vision. Cham: Springer Nature Switzerland, 2022: 205-218.
- [5] Chaki J. Brain Tumor MRI Image Segmentation Using Deep Learning Techniques. Academic Press, 2021.
- [6] Sahayam S. Nenavath R. Jayaraman U. et al. Brain tumor segmentation using a hybrid multi resolution U-Net with residual dual attention and deep supervision on MR images. Biomedical Signal Processing and Control, 2022, 78: 103939.
- [7] Kalaiselvi T. Padmapriya T. Sriramakrishnan P. et al. Development of automatic glioma brain tumor detection system using deep convolutional neural networks. International Journal of Imaging Systems and Technology, 2020, 30(4): 926-938.
- [8] Shareef B. Xian M. Vakanski A. Stan: Small tumor-aware network for breast ultrasound image segmentation. 2020 IEEE 17th international symposium on biomedical imaging (ISBI). IEEE, 2020: 1-5.
- [9] Sriramakrishnan P. Kalaiselvi T. Rajeswaran R. Modified local ternary patterns technique for brain tumour segmentation and volume estimation from MRI multi-sequence scans with GPU CUDA machine. Biocybernetics and Biomedical Engineering, 2019, 39(2): 470-487.
- [10] Nema S. Dudhane A. Murala S. et al. RescueNet: An unpaired GAN for brain tumor segmentation. Biomedical Signal Processing and Control, 2020, 55: 101641.