2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT) | 979-8-3503-8555-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/AINIT61980.2024.10581511

An Optimal U-Net++ Segmentation Method for Dataset BUSI

Jiaqi Shang
Hohai University
Nanjing, China
2220010101@hhu.edu.cn

Yinyi Lai*
Hohai University
Nanjing, China
2120010118@hhu.edu.cn

***Abstract*—The ability to correctly diagnose cancer is crucial for determining prognosis and planning appropriate treatment. At present, there is still a lack of research on various backbone structures within U-Net. This study analyzes the various backbone architectures of U-Net++. By carefully selecting the basic model, the impact of various feature extraction encoders on the results was explored. In order to improve computational efficiency and accuracy, optimization of input size and batch size has been introduced. In addition, by combining transfer learning with pre-trained models, a strategy to improve model performance was proposed: by changing the base model and transfer learning, the base model was ResNet18, the input size was (96,96), and the batch processing was 48. The best U-Net++ model for segmentation on the BUSI dataset was identified using transfer learning as the ImageNet. The structured research methods and conclusions proposed in this study significantly promote the segmentation of breast cancer and provide a promising direction for future research and clinical application.**

Keywords- Segment; U-Net++; Breast cancer; ResNet18

I. INTRODUCTION

Breast cancer is a major global health problem. According to the data of the World Health Organization (WHO) (2020), 2.3 million women will be diagnosed with breast cancer in the world in 2020, including 685000 deaths. By the end of 2020, 7.8 million women have saved their lives after being diagnosed with breast cancer in the past five years[1]. Therefore, it is particularly important to use precise diagnostic techniques to develop effective treatment plans. Due to the flexibility and convenience of ultrasound imaging, it has become a routine method for breast tumors. Accurate segmentation of breast ultrasound (BUS) is a crucial step in diagnosis and treatment planning, but due to the complexity of tumor tissue, including its diverse shapes and unpredictable growth patterns, precise segmentation of cancer tissue faces significant challenges.

At present, medical image segmentation methods can be mainly divided into two categories. The first type is based on traditional segmentation methods. Mainly including threshold based[2], [3], edge detection[4], region growth[5], graph cutting[6], etc. These methods can quickly and accurately segment medical images for relatively simple image segmentation tasks[7]. However, these methods usually require manual selection of features and parameters, which perform poorly in situations of noise and poor image quality. The segmentation effect is also limited for complex anatomical structures and lesion areas with significant changes.

With the development of deep learning[8], powerful learning capabilities have enabled results such as FCN and U-Net to achieve great success in medical image segmentation[9], [10], [11]. For example, Yap et al. systematically evaluated the impact of different FCN variations on breast lesion segmentation and achieved better segmentation results than traditional methods[12]; Haochen Zhao et al. proposed a focal self-attention blockade method for breast ultrasound images. Focusing on the U-Net mechanism will reduce noise and interference in ultrasound images and enhance image segmentation[13]. [14]proposed a dual swin transformer U-Net segmentation framework which aims to incorporate the hierarchical swin transformer into both the encoder and the decoder of the standard U-shaped architecture.

The models that have achieved good segmentation recently are all based on U-Net++[15]. U-Net++ is one of the most advanced models in the current field of medical image segmentation[16]. Its unique architecture, with shrinking paths for feature extraction and broad paths for precise positioning, sets new standards for this field. By exploring various backbone architectures within the U-Net++ framework, this paper effectively improves the accuracy and efficiency of breast cancer segmentation, which provides broad prospects for early detection and treatment strategies.

The contributions of this article are as follows:

(1) The higher accuracy performance of various backbone structures in U-Net is analyzed and evaluated for early prediction of breast cancer.

(2) A high segmentation performance and strong transfer learning U-Net++ architecture have been found in the original U-Net++ structure.

Through detailed experiments and analysis, this research deeply explores the potential of the U-Net++ segmentation model. The basic models are ResNet18 and DenseNet201, and the U-Net++ training results are good. It sets a new benchmark in the accuracy and reliability of breast cancer imaging, paving the way for early detection and strengthening the treatment of breast cancer.

II. METHOD

A. Prophecy

The U-Net++ model is an evolved version of the original U-Net architecture, commonly used to improve segmentation tasks in medical imaging. U-Net++[5]introduces a series of nested and

dense skip connections between the encoder and decoder modules, reducing semantic differences between them and enhancing information flow and gradients through the network. This structure allows for more detailed and accurate feature representation, which is particularly beneficial for segmenting complex medical images and ensuring higher accuracy and reliability in medical diagnosis. As shown in Figure 1, U-Net++ consists of U-Nets of different depths, and its decoders are densely connected at the same resolution through redesigned skip connections. Compared with the classic U-Net architecture, the significant improvement of U-Net++ is attributed to the advantages provided by the redesigned skip connections and extended decoders, which together enable image features to be gradually aggregated in the network both horizontally and vertically.

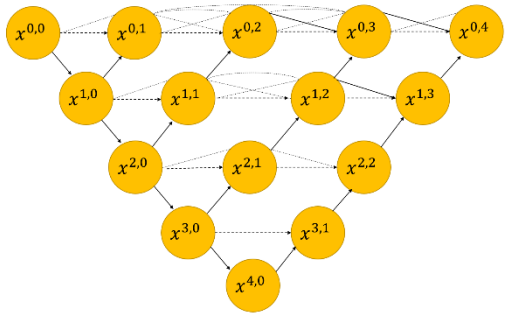

Figure 1. U-Net++ Model Structure.

B. Foundation

Based on U-Net's success, U-Net++ introduces an enhanced architecture with nested dense skip paths, aiming to further improve segmentation accuracy. It obtains image feature information through an encoder and adds skip connections during the decoding process to preserve the image feature information, but is still constrained by the inherent limitations of convolution. The derivation of its formula can be expressed as:

Encoder: The encoder path of U-Net++ is similar to the original U-Net model, using convolutional and pooling layers to gradually reduce the size of the feature map while increasing the depth of the features. If the input feature map of layer i is X_i, then the output feature map of layer i+1, X_{i+1}, can be represented as:

$$X_{i+1}=\text{MaxPool}(\text{ReLU}(\text{Conv}(X_i)))$$

Among them, Conv represents convolution operation, ReLU represents activation function, and MaxPool represents max pooling operation.

Decoder: U-Net++'s decoder path adopts nested skip connections, allowing the feature maps from the encoder path to be multi-scale fused with the feature maps from the decoder path. If the decoder feature map of the i-th layer is Y_{i-1}, then the decoder feature map Y_{i-1} of the i-1st layer can be represented as:

$$Y_{i-1}=\text{ConvTranspose}(\text{Concat}(Y_i, X_{i-1}, Y_{i-1,1}, Y_{i-1,2}, \ldots, Y_{i-1,i-1}))$$

Among them, ConvTranspose represents the transpose convolution operation, Concat represents the concatenation operation of feature maps, and $Y_{i-1,j}$ represents the j^{th} nested skip connected feature map in the i-1st layer decoder path.

Output layer: U-Net++ uses a convolutional layer to map the final feature map to the number of target categories and obtain the final segmentation result. If the output feature map is set to Y_0, the segmentation result S can be represented as:

$$S=\text{Conv}(Y_0)$$

Among them, the output feature map Y_0 is the top-level feature map of the decoder path.

C. Methodology

Figure 2 shows the method of breast cancer detection. The steps are as follows:

Step 1: Image acquisition: Select the dataset, including images and models.

Step 2: Image preprocessing: Adjust the image to a consistent and appropriate size. The purpose of adjusting the image size is to segment the image more efficiently and improve the efficiency of breast cancer detection.

Step 3: Feature extraction and image segmentation: This is accomplished by a U-Net++ model with dense continuous skip connections.

Step 4: Detection: test the U-Net++ model on the sample image to obtain the corresponding image information, display the segmentation area in the image, and facilitate the detection of breast cancer.

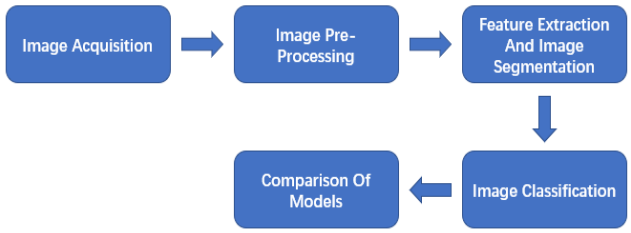

Figure 2. Methodological Steps.

This article uses different backbone architectures such as ResNet, VGG, and DenseNet to provide unique advantages for feature extraction in the U-Net++ model. ResNet is renowned for its deep residual learning framework, which can train deeper networks. The characteristics of VGG are simplicity and depth, using small convolutional filters for feature extraction. DenseNet improves this by connecting each layer to each other layer in a feedforward manner, ensuring maximum information flow between layers in the network. We utilize the different characteristics of these architectures to enhance the segmentation effect of U-Net++.

In the experiment, input shape and batch size are key parameters. The input shape determines the size of the input model's image, thereby affecting the way the model perceives and processes image features. The batch size affects the learning dynamics of the model, with larger batches providing more stable gradient estimates and smaller batches providing more frequent model updates.

The evaluation indicators are evaluated using the average dice coefficient and loss indicator. Quantify performance using the average dice coefficient, which measures the similarity

between predicted segmentation and actual segmentation, where 1 represents complete overlap and 0 represents no overlap.

III. EXPERIMENT

A. Dataset

The dataset we use, BUSI, includes ultrasound breast images of women aged 25 to 75. The number of patients is 600 female patients. This sample was collected in 2018 using the LOGIQ E9 ultrasound system and LOGIQ E9 Agile ultrasound from the Baheya Women's Cancer Early Classification and Treatment Hospital in Cairo (Egypt). Divide the data into three categories: normal, benign, and malignant. This dataset consists of 780 images from different women, with an average image size of 500x500 pixels. Among 780 tumor images, 133 are normal images of non-cancerous masses, 437 are images of cancerous masses, and 210 are images of benign masses. The dataset BUSI can be used for online research[17]. We will use 75% of the images and masks as training data, and the remaining 25% of the images and data as testing data. The partial data display is shown in Figure 3.

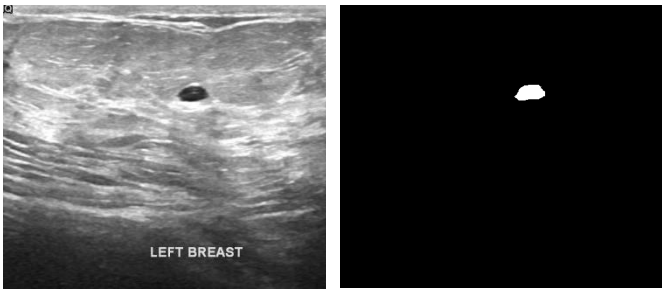

Figure 3. Dataset Images.

B. Implementation Details

The software environment is the AutoDL server rental platform, and the operating system is Linux. The hardware environment is: the computing power model is RTX 4090, the graphics memory is 24GB, and the CPU model is 12 core Xeon (R) Platinum 8352V.

According to the above experimental settings, the number of cycles is generally limited to 50 or less. If additional time is needed, unnecessary cycles can be reduced. When conducting experiments on the dataset, Mean Dice was used in the loss function. Additionally, we use flipping, translation and other operations to enhance the data.

C. Comparison of Basic Models

Through the previous knowledge about the encoder part, we know that in the U-Net++ model, the encoder part is crucial and is mainly responsible for extracting the features of the image and providing necessary information for subsequent image segmentation tasks.

In this experiment, we compare the performance of different base models in the encoder on the segmentation task, including the ResNet series (ResNet18, ResNet34, ResNet50, ResNet152), ResNeXt series (ResNeXt101_12x4d, ResNeXt101_12x8d), DenseNet (DenseNet161), and VGG series (VGG13, VGG16). These models all perform well in image classification tasks, but their performance varies in segmentation tasks.

The experimental results in Table 1 show that ResNet18 performs the best among these basic models, reaching a Dice coefficient of 0.672. This may be because ResNet18 has fewer parameters and is easier to train compared to other deeper networks, while also maintaining sufficient feature extraction capabilities. Deeper networks such as ResNet152, although theoretically capable of extracting more complex features, may experience performance degradation in practical applications due to overfitting or insufficient training. In addition, the VGG series performed poorly in this experiment due to its relatively simple structure. Based on the experimental results, ResNet18 was selected as the basic model for subsequent experiments, because it showed good performance and high efficiency in breast cancer segmentation tasks.

Table 1 The impact of input image size

Serial Number	*Base Model*	*Input Shape*	*Batch Size*	*Eval Dice*
1	**ResNet18**	**(128,128)**	**8**	**0.672**
2	ResNet34	（128,128）	8	0.651
3	ResNet50	（128,128）	8	0.589
4	ResNet152	（128,128）	8	0.647
5	ResNeXt101_32x4d	（128,128）	8	0.596
6	ResNeXt101_32x8d	（128,128）	8	0.621
7	DenseNet161	（128,128）	8	0.626
8	VGG13	（128,128）	8	0.544
9	VGG16	（128,128）	8	0.569

D. Input Image Size Comparison

The aim of this experiment is to explore the impact of different input sizes on model performance, in order to determine the optimal input size. In the experiment, ResNet18 was used as the basic model to test various input sizes, including (32x32), (64x64), (96x96), (128x128), (192x192), and (256x256).

Table 2 The impact of input image size

Serial Number	*Base Model*	*Input Shape*	*Batch Size*	*Eval Dice*
1	**ResNet18**	**(96,96)**	**8**	**0.690**
2	ResNet18	(32,32)	8	0.685
3	ResNet18	(64,64)	8	0.686
4	ResNet18	(128,128)	8	0.672
5	ResNet18	(192,192)	8	0.625
6	ResNet18	(256,256)	8	0.615

As shown in Table 2, the experimental results show that when the input size is (96x96) and the input size is 96x96, the model reaches the highest Dice coefficient of 0.690. This indicates that in this specific segmentation task, a moderate input size can provide sufficient detailed information while avoiding

the computational burden and overfitting risk caused by excessive size. An input size that is too small, such as 32x32, may lose too much detail, while a size that is too large, such as 256x256, may make it difficult for the model to capture local features. Based on the experimental results, subsequent experiments will use (96x96) as the input size to balance model performance and computational efficiency.

E. The Effect of Batch Size

In this experiment, we investigated the impact of different batch sizes on model performance under fixed input size (96x96) and base model (ResNet18). Multiple batch sizes ranging from 8 to 128 were tested in the experiment, including 8, 32, 48, 64, and 128.

Table 3 The impact of batch size

Serial Number	*Base Model*	*Input Shape*	*Batch Size*	*Eval Dice*
1	**ResNet18**	**(96,96)**	**48**	**0.696**
2	ResNet18	(96,96)	8	0.663
3	ResNet18	(96,96)	32	0.662
4	ResNet18	(96,96)	64	0.659
5	ResNet18	(96,96)	128	0.627

As shown in Table 3, the experimental results show that when the batch size is 48, the model reaches the highest Dice coefficient of 0.696. This may be because a moderate batch size can provide sufficient sample information while maintaining good gradient estimation and memory efficiency. Based on the experimental results, subsequent experiments will use a batch size of 48 to achieve the best model performance.

F. The Impact of Transfer Learning and Segmentation Algorithms

This experiment aims to explore the impact of transfer learning and different segmentation algorithms on model performance. ResNet18 was used as the basic model in the experiment, with an input size of (96x96) and a batch size of 48. Different segmentation algorithms (U-Net, U-Net++, FPN) and whether pre-training weights (ImageNet) were used were tested.

Table 4 The impact of transfer learning and segmentation algorithms

Serial Number	*Transfer Learning*	*Segmentation Algorithm*	*Eval Dice*
1	None	smp.UnetPlusPlus()	0.696
2	None	smp.Unet()	0.656
3	None	smp.FPN()	0.664
4	**ImageNet**	**smp.UnetPlusPlus()**	**0.697**

As shown in Table 4, the experimental results show that when using the U-Net++ algorithm and pre-trained weights (ImageNet), the model achieved the highest Dice coefficient of 0.697. This may be because the U-Net++ algorithm can capture more contextual information and detailed features through nested skip connections and deep supervision, thereby improving the accuracy of segmentation. Meanwhile, using pre-trained weights can accelerate the convergence speed of the model and improve its generalization ability. We believe that ResNet18 effectively alleviates the problem of gradient vanishing through residual connections while maintaining a low number of parameters and computational complexity, enabling the model to better learn the features of segmentation tasks.

IV. CONCLUSIONS

Through this study, we found that in the comparison of basic models, ResNet18 effectively learned the features of segmentation tasks through residual connections while maintaining low computational complexity, demonstrating the best performance. Secondly, different input sizes have a significant impact on model performance, and an appropriate input size can strike a balance between information preservation and computational efficiency, thereby improving model performance. At the same time, larger batch sizes help provide more stable gradient estimates, which helps the model converge more effectively. In addition, transfer learning and the use of more advanced segmentation algorithms such as U-Net++ can further improve the performance of the model, which may be due to the better initialization provided by pre-trained weights, while U-Net++ can more effectively capture multi-scale information by adding nested connections and deep supervision mechanisms. Overall, by selecting the appropriate base model, input size, batch size, and segmentation algorithm, we can effectively improve the performance of segmentation tasks. Our experimental results provide valuable references for all researchers to choose appropriate models and parameters in practical applications.

The main challenges currently faced include the complexity of U-Net++ models, overfitting risks, hyperparameter tuning, and so on. With the continuous development of deep learning, in the future, we will try more basic models and more possibilities. We will continue to explore the integration of attention mechanisms and generative models such as GAN to carry out fine segmentation in breast cancer imaging. Meanwhile, we will also attempt data augmentation techniques and integrate more types of imaging data to enhance the robustness and generalization of the model. In addition, exploring the interpretability of model decisions can provide valuable insights for clinical doctors and improve trust and adoption rates in the clinical environment.

Our future work will focus on integrating advanced attention mechanisms and using GANs for data augmentation to enhance robustness. Combining ultrasound images with other modalities like MRI can improve diagnostic accuracy. Developing lightweight, real-time U-Net++ models for portable devices will facilitate clinical adoption. Enhancing model interpretability with techniques like saliency maps will build clinical trust. Additionally, hyperparameter optimization and cross-dataset validation will ensure robustness and generalizability, advancing breast ultrasound image segmentation and improving early detection and treatment of breast cancer.

REFERENCES

[1] D. Rajput and B. J. Bejoy, "State-of-art Techniques for Classification of Breast Cancer: A Review," in 2022 5th International Conference on Contemporary Computing and Informatics (IC3I), Dec. 2022, pp. 433–437. doi: 10.1109/IC3I56241.2022.10073435.

[2] B. Poorani and R. Khilar, "Identification of Polycystic Ovary Syndrome in ultrasound images of Ovaries using Distinct Threshold based Image Segmentation," in 2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT), Gharuan, India: IEEE, May 2023, pp. 570–575. doi: 10.1109/InCACCT57535.2023.10141800.

[3] J. Zeng and Y. Li, "Knee Joint Magnetic Resonance Images Recognition Based on Iterative Threshold Segmentation," in 2022 2nd International Conference on Computer Graphics, Image and Virtualization (ICCGIV), Chongqing, China: IEEE, Sep. 2022, pp. 139–144. doi: 10.1109/ICCGIV57403.2022.00034.

[4] M. Kim and B.-D. Lee, "A Simple Generic Method for Effective Boundary Extraction in Medical Image Segmentation," IEEE Access, vol. 9, pp. 103875–103884, 2021, doi: 10.1109/ACCESS.2021.3099936.

[5] K. Zhang, F. Wu, J. Sun, G. Yang, H. Shu, and Y. Kong, "Iterative Seeded Region Growing for Brain Tissue Segmentation," in 2022 IEEE International Conference on Image Processing (ICIP), Bordeaux, France: IEEE, Oct. 2022, pp. 886–890. doi: 10.1109/ICIP46576.2022.9897303.

[6] J. Sun, Y. Zhang, J. Zhu, J. Wu, and Y. Kong, "Semi-Supervised Medical Image Semantic Segmentation with Multi-scale Graph Cut Loss," in 2021 IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA: IEEE, Sep. 2021, pp. 624–628. doi: 10.1109/ICIP42928.2021.9506098.

[7] F. H. H. Shajin, B. A. Devi, N. B. Prakash, G. R. Sreekanth, and P. Rajesh, "Sailfish optimizer with Levy flight, chaotic and opposition-based multi-level thresholding for medical image segmentation," Soft Comput., vol. 27, no. 17, pp. 12457–12482, Sep. 2023, doi: 10.1007/s00500-023-07891-w.

[8] X. Liu, L. Song, S. Liu, and Y. Zhang, "A Review of Deep-Learning-Based Medical Image Segmentation Methods," Sustainability, vol. 13, no. 3, p. 1224, Jan. 2021, doi: 10.3390/su13031224.

[9] Y. Wang et al., "Deep Attentive Features for Prostate Segmentation in 3D Transrectal Ultrasound," IEEE Transactions on Medical Imaging, vol. 38, no. 12, pp. 2768–2778, Dec. 2019, doi: 10.1109/TMI.2019.2913184.

[10] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation From CT Volumes," IEEE Trans. Med. Imaging, vol. 37, no. 12, pp. 2663–2674, Dec. 2018, doi: 10.1109/TMI.2018.2845918.

[11] S.-T. Tran, C.-H. Cheng, T.-T. Nguyen, M.-H. Le, and D.-G. Liu, "TMD-Unet: Triple-Unet with Multi-Scale Input Features and Dense Skip Connection for Medical Image Segmentation," Healthcare, vol. 9, no. 1, p. 54, Jan. 2021, doi: 10.3390/healthcare9010054.

[12] M. H. Yap et al., "Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks," IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 4, pp. 1218–1226, Jul. 2018, doi: 10.1109/JBHI.2017.2731873.

[13] H. Zhao, J. Niu, H. Meng, Y. Wang, Q. Li, and Z. Yu, "Focal U-Net: A Focal Self-attention based U-Net for Breast Lesion Segmentation in Ultrasound Images," in 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Jul. 2022, pp. 1506–1511. doi: 10.1109/EMBC48229.2022.9870824.

[14] A. Lin, B. Chen, J. Xu, Z. Zhang, G. Lu, and D. Zhang, "DS-TransUNet: Dual Swin Transformer U-Net for Medical Image Segmentation," IEEE Trans. Instrum. Meas., vol. 71, p. 4005615, 2022, doi: 10.1109/TIM.2022.3178991.

[15] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet plus plus: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation," IEEE Trans. Med. Imaging, vol. 39, no. 6, pp. 1856–1867, Jun. 2020, doi: 10.1109/TMI.2019.2959609.

[16] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications," IEEE Access, vol. 9, pp. 82031–82057, 2021, doi: 10.1109/ACCESS.2021.3086020.

[17] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," Data in Brief, vol. 28, p. 104863, 2020, doi: https://doi.org/10.1016/j.dib.2019.104863.