



Application of U-Net and its variants in ultrasound image segmentation

Yuxiang Wang¹, Miao Zhou², Fangfang Chen¹, Jintao Duan¹, Liangqing Lin³, Qinghua Wu³, Wenhui Guo⁴, Haipo Cui¹

¹School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China. ²Jiangsu Cancer Hospital, Nanjing 213164, China. ³Anesthesiology, The First Hospital of Putian, Putian 351100, China. ⁴School of Anesthesiology, Second Military Medical University/Naval Medical University, Shanghai 200433, China.

Corresponding author: Haipo Cui.

Declaration of conflict of interest: None.

Received September 27, 2024; Accepted January 22, 2025; Published April 1, 2025

Highlights

- This review provides a comprehensive overview of various U-Net architectures and their variants, including the original U-Net, U-Net++, Attention U-Net, and ResU-Net, along with a discussion on potential improvements to the U-Net architecture.
- The differences between these network architectures are analyzed in terms of training complexity and computational requirements.
- The review delves into the application of U-Net and its variants in ultrasound imaging, discussing both the advantages and limitations of each model in various ultrasound contexts. Relevant literature on the application of each network architecture in ultrasound is also summarized.

Abstract

Ultrasonography plays an important role in the fields of obstetrics, gynecology, cardiology, and hepatology, as well as ultrasound-guided nerve blocks, interventional therapy, and surgical navigation due to its non-invasive, real-time imaging and radiation-free characteristics. Recently, with the advancement of artificial intelligence, machine learning and deep learning algorithms have brought significant innovations to ultrasound imaging technology in the medical field. U-Net is widely recognized as one of the most commonly used deep learning models in medical image processing. This paper explores the application of the U-Net family of models in ultrasound imaging. The network architecture of the original U-Net, comprising encoder and decoder components, is first delineated. Next, classical variants, such as U-Net++, Attention U-Net, and ResU-Net, are introduced. The application of U-Net models in ultrasound and their segmentation performance are then reviewed, with Dice coefficients highlighted as the primary evaluation metric. Finally, the paper provides a comparative analysis of the advantages and disadvantages of the U-Net family of models.

Keywords: U-Net, U-Net++, Attention U-Net, ResU-Net, ultrasonography, segmentation performance

1 Introduction

Ultrasound, a widely utilized medical diagnostic tool, possesses real-time imaging capabilities, facilitating the real-time visualization of nerves, blood vessels, and muscles. This real-time imaging capability minimizes waiting times for both patients and medical practitioners. Despite its widespread use, ultrasound equipment has inherent limitations. The low contrast in

ultrasound images often results in similar echo characteristics across different tissues, posing challenges for doctors in distinguishing specific structures. For instance, during a lumbar plexus nerve block, the boundaries of the cloverleaf structure, which includes the quadratus lumborum, major lumbar muscle, and erector spinae muscle, may be unclear, leading to potential misidentification of abdominal cavity as transverse processes. Furthermore, the quality of



Table 1. Summary of the applicability, advantages, and disadvantages of each network structure

Model	Structural characteristics	Advantage	Disadvantage	Applicability
U-Net	Infrastructure	Aids in capturing features effectively.	May have limitations in dealing with multi-scale features.	Widely used in segmentation tasks for ultrasound images
U-Net++	Jump connections and deep supervision are introduced	Multi-level feature fusion contributes to comprehensive context understanding.	The network structure is relatively more complex.	More suitable for ultrasound images with multi-scale features
Attention U-Net	Attention mechanisms were introduced	The introduction of an attention mechanism improves the ability to capture key features.	More training samples are needed to learn the attention weights efficiently.	For tasks that require better attention to important structures
ResU-Net	Residual learning and U-Net were combined	Combines the advantages of residual learning and U-Net to provide deeper feature learning.	More training samples are needed to train the deep network efficiently.	For tasks requiring deeper feature learning and better generalization performance

ultrasound images is influenced by various positions and angles of the ultrasound probe [1]. This variability can result in images of differing quality and readability. As a result, diagnosing and testing diseases rely heavily on the subjective experience of the physician.

In recent years, the integration of artificial intelligence, particularly machine learning and deep learning, has gained prominence in the medical field [2]. Machine learning algorithms applied to ultrasound enable automated image analysis for tasks such as lesion detection, disease screening, segmentation, and feature extraction [3]. This automation improves the accuracy of disease detection and early diagnosis. Specifically, in breast tumor detection, ultrasound image segmentation plays a critical role. Deep learning models, particularly those based on the U-Net architecture, have proven effective for detecting and segmenting breast tumors [4-7].

The U-Net structure, recognized as one of the most widely employed deep learning models in image processing, excels in image segmentation and feature learning. Its proficiency in localizing and segmenting cells, tissues, or lesions in medical images makes it a powerful tool for precise segmentation tasks in medical images [8-11]. Ultrasound images, characterized by low contrast and high noise, present challenges for traditional image processing methods such as thresholding, edge detection, and mathematical morphology. The U-Net architecture addresses these challenges by employing jump connections, which effectively preserve high-resolution information and prevent information loss. This adaptability makes U-Net particularly well-suited for processing intricate medical ultrasound images. This paper focuses on the application of the U-Net family of models in ultrasound imaging.

II Methods and materials

Performance comparison with different methods

Each network structure has its unique characteristics, and no single model is universally applicable. In practical scenarios, experiments and evaluations should be tailored to the specific task, dataset features, and available computational resources. **Table 1** summarizes the applicability, advantages, and disadvantages of each network structure. In addition, the differences in training complexity and computational requirements across the various models are summarized in **Table 2**.

U-net

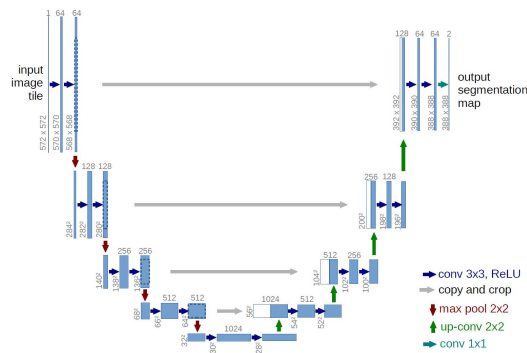
The U-Net architecture was proposed in 2015 by German researchers Olaf Ronneberger, Philipp Fischer, and Thomas Brox. The name “U-Net” comes from the network’s “U”-shaped structure, which is a distinctive feature of this architecture [12]. As depicted in **Figure 1**, the architecture comprises an encoder and a decoder. The encoder progressively extracts feature maps through multiple convolution and pooling operations, while the decoder enlarges the resolution through up-sampling. Skip connections link to the encoder and decoder, facilitating the fusion of feature maps, ultimately producing the segmentation result at the output layer.

Encoder

The encoder plays a crucial role in the U-Net architecture by extracting features from the input image through a sequence of convolutions, rectified linear unit (ReLU) activations, and max pooling. After each convolution, a feature map

Table 2. The differences in training complexity and computational requirements across the various models

Model	Training complexity	Computational requirements
U-Net	Due to its simple structure, U-Net requires fewer parameters during the training process, which makes the training process more efficient.	It has a relatively low computational load.
U-Net++	Due to the structural complexity, the number of parameters during training is higher than that of U-Net, necessitating greater computational resources for both training and tuning.	The computational requirements of U-Net++ have increased substantially due to the incorporation of additional skip and dense connections.
Attention U-Net	Although the attention mechanism is incorporated, the number of parameters in Attention U-Net does not increase as significantly as in U-Net++, while maintaining training efficiency.	The incorporation of the attention mechanism introduces additional computational and memory overhead.
ResU-Net	Despite the introduction of residual connectivity to accelerate training convergence, the structure of the residual network results in a higher training complexity compared to that of a regular U-Net.	Each convolutional layer incorporates residual connections, meaning that two paths (standard and residual) are utilized in each layer. This addition increases the computational burden, particularly when deep networks are involved, further exacerbating computational complexity.

**Figure 1. U-Net network framework.** This figure is cited from [12].

is produced. The ultimate downsampling layer comprises 1024 feature maps, each of size 28×28. The following sections explain the convolution operation in detail.

An input image of dimensions $H \times W \times C$ is processed, where H represents the height, W is the width, and C is the number of channels. A convolution operation is then performed, where each convolutional layer contains numerous convolutional kernels to execute the convolution operation on the input image. In the original U-Net structure, the convolution kernel is 3×3, with a stride of 1 and no padding. Upon completion of the convolution, a feature map is generated with the same number of convolution kernels, and the size of the feature map is as follows:

$$\text{Output Size} = \frac{\text{Input Size} - \text{Kernel Size} + 2 \times \text{Padding}}{\text{stride}} + 1$$

After the convolution, ReLU activation function is applied for feature extraction. ReLU is commonly used to enhance the network's ability to learn complex data patterns by introducing non-linearity [13]. It is computationally simple,

has a straightforward derivative (constant at 1 in most regions and 0 in others), and helps alleviate the vanishing gradient problem [14]. Its definition is as follows:

$$g(t) = \max(0, t)$$

In U-Net, ReLU improves the nonlinear representation of the network, enabling it to capture complex feature mappings and enhancing performance in medical image segmentation tasks. Each convolutional layer in the original U-Net undergoes two convolutions with convolutional kernels, and after each convolution, ReLU function is applied for feature extraction.

Following this, max pooling is typically employed in the pooling layer of U-Net. Max pooling involves sliding a window (of size $n \times n$) across the feature map with a stride of t and selecting the maximum value within the window at each position. The formula for calculating the maximum value is:

$$\text{Max_Value} = \max(X_{ij}, X_{i+1,j}, X_{i,j+1}, X_{i+1,j+1})$$

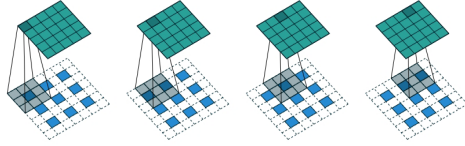
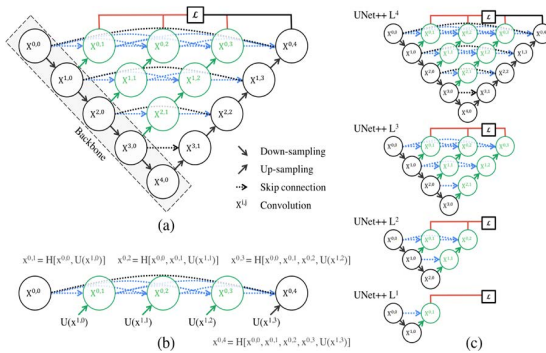
Where X_{ij} denotes the value at coordinate (i, j) of the input feature map.

Decoder

The decoder, functioning as a feature fusion network, aims to progressively restore the image dimensions and reconstruct the spatial details of the input image. This is achieved by integrating the feature maps from corresponding layers of the encoder with those of the decoder. Each stage of the decoder involves an upsampling operation, a convolution operation, and the amalgamation of feature maps.

Table 3. Summary of the development/improvements in the upsampling procedure

Presenters	Year	Achievement
Goodfellow et al. [16]	2014	GANs typically use transposed convolutional layers for upsampling.
Long et al. [17]	2015	Fractionally strided convolution was introduced for the first time.
Noh et al. [18]	2015	Applied to semantic segmentation tasks.

**Figure 2. Transposed convolution example.** This figure is cited from [15].**Figure 3. U-Net++ network framework.** (a) UNet++ consists of an encoder and decoder that are connected through a series of nested dense convolutional blocks. The main idea behind UNet++ is to bridge the semantic gap between the feature maps of the encoder and decoder prior to fusion. For example, the semantic gap between $(X^{0,0}, X^{1,3})$ is bridged using a dense convolution block with three convolution layers. In the graphical abstract, black indicates the original U-Net, green and blue show dense convolution blocks on the skip pathways, and red indicates deep supervision. Red, green, and blue components distinguish UNet++ from U-Net; (b) Detailed analysis of the first skip pathway of UNet++; (c) UNet++ can be pruned at inference time, if trained with deep supervision. This figure is cited from [21].

The decoder begins with an upsampling operation, typically achieved through transposed convolution utilizing transposed convolution (also called inverse convolution). Unlike conventional convolution, transposed convolution accomplishes upsampling by inserting zeros between the outputs. A specific example of this process is illustrated in **Figure 2**. Suppose the input is a 5×5 matrix. Using a 3×3 convolution kernel with a stride of 2×2 and 1×1 zero-padding results in a 7×7 output matrix [15].

After up-sampling, the dimensions of the resulting feature maps are adjusted to match those of the feature maps from the corresponding encoder layer. **Table 3** summarizes the develop-

ment/improvements in the upsampling procedure.

Since transposed convolution enlarges the image rather than restoring it to its original size, feature concatenation is subsequently performed. This step is typically executed in a 'copy and crop' manner: the upsampled feature maps are copied to the corresponding encoder layer, and the encoder feature maps are cropped to match the dimensions. Typically, the cropping operation preserves the central region, ensuring identical dimensions in both height and width for both feature maps. Ultimately, the cropped feature maps are concatenated with the feature maps from the corresponding encoder level. This step facilitates the integration of low- and high-level features, which enhances the network's ability to recover image features.

Finally, a convolution operation is applied to the fused feature maps. This convolution in the decoder follows the same approach as in the encoder. After the first layer of the decoder completes its operations, this process is repeated for subsequent layers. Similar to the encoder, each decoder layer undergoes two convolution operations, with the image size doubling with each layer. However, it is noteworthy that in the final convolution of the last layer in the decoder, the original U-Net structure uses a 1×1 convolution kernel and does not apply an activation function.

Mainstream U-Net variants

U-Net++

U-Net++ was originally proposed by researchers from the University of Belgrade in Serbia, to address some limitations of the original U-Net architecture, such as information bottlenecks and feature loss in image segmentation tasks. U-Net++ introduces two key improvements over the traditional U-Net. First, it reconfigures the jump paths by incorporating jump connections. Unlike traditional U-Net networks, which solely fuses feature maps from encoders and decoders at the same level, U-Net++ introduces a dense convolutional block. This block facilitates jump connections, allowing the fusion of feature maps from encoders and decoders at diverse levels [19]. In this context, the formulation of the jump-connect path is delineated as

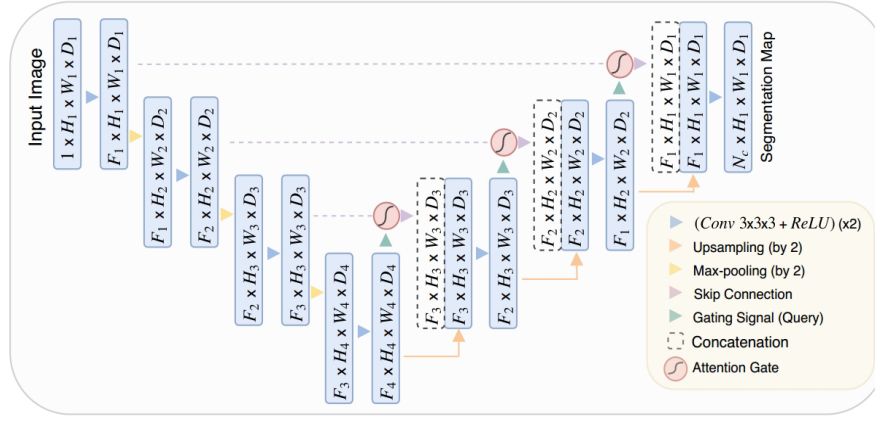


Figure 4. Attention U-Net network framework. This figure is cited from [24].

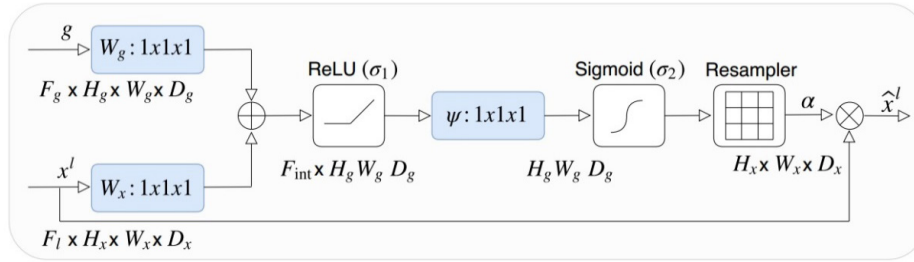


Figure 5. Schematic diagram of attention gate. This figure is cited from [24].

follows:

$$x^{i,j} = \begin{cases} R(n) & \text{If } n = 0 \\ R([x^{m,h}]_{h=0}^{m-1}, t(x^{m+1,n-1})), & \text{If } n > 0 \end{cases}$$

Second, deep supervision is incorporated in U-Net++, which involves integrating supervisory signals (loss functions) at various levels in the network [20]. These signals are introduced at each jump connection, enhancing the network's ability to learn features at various levels, thus refining segmentation performance. **Figure 3** illustrates a schematic representation of this network.

Attention U-Net

Attention U-Net introduces an attention mechanism that allows the network to automatically focus on specific areas of an image, mimicking human visual perception. When humans perceive stimuli, their attention is directed to particular areas based on their needs. This attention mechanism has been adopted in deep learning to enhance model performance by dynamically adjusting the focus on different parts of the input data, improving both performance and generalization [22]. Specifically, through the dynamic adjustment of attention weights, attention mechanisms enable the model to process different parts of input data in a more targeted manner.

Various types of attention mechanisms, such as channel attention and temporal attention, are tailored for various applications and tasks [23]. In U-Net architectures, the introduction of skip connections allows the concatenation of encoder features with corresponding decoder layers. However, this can lead to feature redundancy. To address this issue, Oktay et al. proposed the integration of an Attention Gate into the U-Net architecture, as depicted in **Figure 4** [24]. The integration of the Attention Gate within the skip connection dynamically adjusts the degree of attention on features, mitigating redundancy and enhancing the overall performance and generalization ability of the network.

In the design of attention gates, two inputs are initially involved, namely x and g . x denotes the feature map of the preceding layer in the encoder hierarchy, and g represents the global features, typically acquired by performing global pooling on the encoder output. Following this, x and g undergo a 1×1 convolution operation, followed by element-wise summation to derive the weighted, fused features. Subsequently, a non-linear transformation is introduced via the ReLU activation function, followed by another 1×1 convolution to enhance feature characterization. Attention weighting coefficients are then generated using the Sigmoid function, reflecting the importance of various features. Finally, the resulting coefficients are used to resample the features using trilinear interpolation and

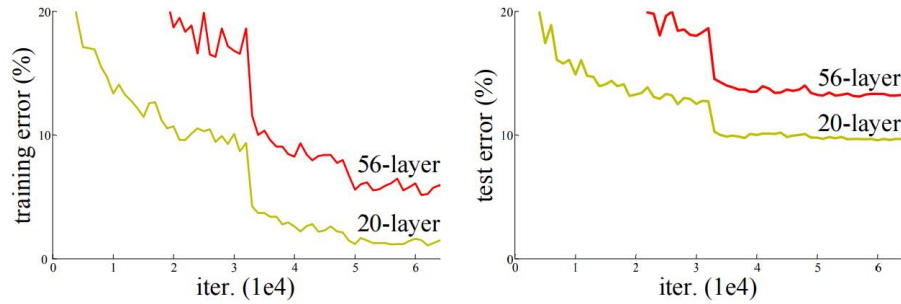


Figure 6. Training error and test error. This figure is cited from [25].

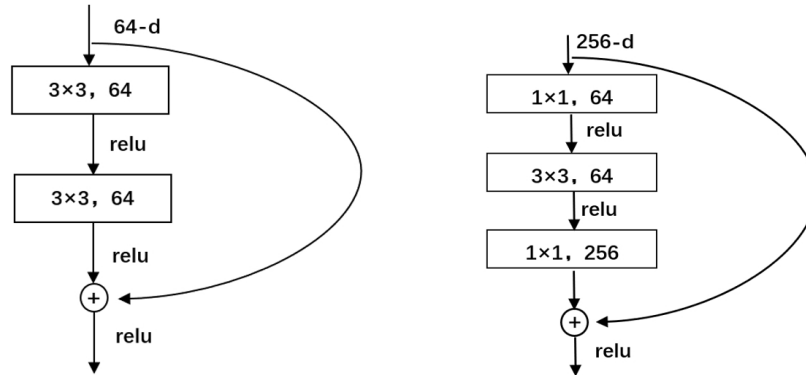


Figure 7. Residual structure. This figure is cited from [25].

multiply them with x to produce the final output, which is used for segmentation. This entire process is depicted in **Figure 5**.

ResU-Net

The design inspiration for ResU-Net is mainly derived from Residual Network and U-Net. Residual Network is a highly successful deep neural network structure that addresses the issues of gradient vanishing and gradient explosion during training by introducing residual connections. This innovation allows for the successful training of deeper network models.

As shown in **Figure 6**, the training and testing errors gradually increase as the network deepens. When a network is made deeper by adding more layers, issues of gradient vanishing or gradient explosion arise, hindering the model's convergence. Moreover, as the network continues to converge, it faces the problem of degradation, where the network's performance rapidly declines once a certain level of accuracy is achieved. To address these issues, a residual structure is introduced. The residual structure consists of two key components: the Primary Path and Shortcut Connection. As depicted in **Figure 7**, the left side represents the residual structure for networks with fewer layers, while the right side represents the residual structure employed in deeper networks [21].

In the residual structure, the primary pathway includes a convolutional layer followed by the ReLU activation function. The Shortcut Connection links the input directly to the output, bypassing the convolutional layer. This connection allows the input feature matrix to be added to the output feature matrix after it has passed through the primary pathway. Importantly, the feature matrix after passing through the Shortcut Connection must have the same dimensions as the output feature matrix of the primary pathway. The ResU-Net architecture integrates this residual connection structure into the standard U-Net framework. **Figure 8** delineates its configuration.

Application in ultrasound

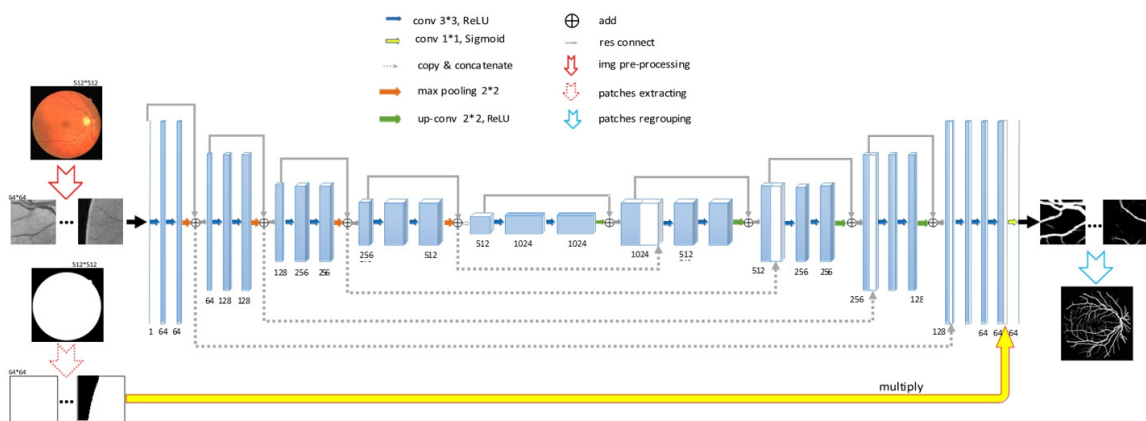
U-Net

Upon its proposal, U-Net has become a classical neural network in image segmentation, adept at extracting and processing distinct image features through the acquisition of feature representations. This capability has led to its widespread adoption in ultrasound imaging, with both the original U-Net architecture and its refinements being employed. Initially designed for the analysis of microscopic images, such as cell and tissue sample recognition, U-Net has since found significant applications in ul-

Table 4. The applications of U-Net in ultrasound image analysis

Presenters	Analytical objectives	Results
Zyuzin et al. [29]	Left ventricle endocardial border	Accuracy of LV segmentation is up to 92.3%
Wang et al. [30]	Brachial plexus	Dice coefficient got 0.6490
Tian et al. [31]	Brachial plexus	Accuracy with an IoU of 68.50%
Li et al. [32]	Levator hiatus	Gets a mean Dice of 96.4%±0.7%
Díaz-Vargas et al. [33]	Peripheral nerve	Average Dice coefficient of 0.70
Zou et al. [34]	Ovarian lesions	The AUC was 0.94, and the accuracy was 0.89.
Yang et al. [35]	Vascular lumen and medium-adventitia	0.87 JM, 0.82 mm HD for lumen and 0.86 JM, 1.07 mm HD for media (40 MHz dataset)
Ben et al. [36]	Breast cancer	Dice coefficient, Precision and Recall all reached over 80%.
Shodiq et al. [37]	Intravascular deep vein thrombosis segmentation	The mean IoU is 84.9% accurate, the measure of the mean Hausdorff distance is 4.17 of the score.
Suissa et al. [38]	Transversus abdominis plane	Global Dice score of 73.31%
Oghli et al. [39]	Automatic measurement of BPD and HC by segmentation of fetal head	Achieved a DSC of 0.95 and HD of 4.5
Lin et al. [40]	Laparoscopic ovarian	The ultrasound image showed no unnecessary interference noise, and the image clarity was significantly improved.

Note: LV, left ventricular; IoU, Intersection over Union; AUC, Area Under the Curve; JM, Jaccard Measure; HD, Hausdorff Distance; HC, Head circumference; BPD, Biparietal diameter; DSC, Dice similarity coefficient.

**Figure 8. ResU-Net network framework.** This figure is cited from [26].

trasound imaging. These applications include the segmentation of organs and tissues, the detection and segmentation of lesions, and cross-modal image registration. The following sections provide detailed exploration of these applications.

Organ and tissue segmentation

Ultrasound imaging, including liver and cardiac ultrasound, has gained popularity due to its real-time imaging capabilities, cost-effectiveness, and non-invasive nature, especially when compared to CT scans. However, ultrasound images tend to have high noise, which complicates the accurate identification of structures. Deep learning algorithms, notably U-Net, have proven effective in overcoming these challenges, enabling the segmentation and recognition of organs and tissues in real-time. This capa-

bility allows ultrasound machines to display the boundaries of these structures clearly, helping physicians detect them more efficiently and improving diagnostic and treatment efficiency.

Lesion detection and segmentation

Cancer, characterized by high morbidity and mortality, profoundly impacts individuals both psychologically and economically. Liver cancer, ranking as the second most common cancer after lung cancer, accounts for 11% of global cancer-related mortality [27]. Additionally, breast cancer remains a significant concern, making early screening essential to improve outcomes. Compared to other screening methods, ultrasound is less harmful to patients; however, manual segmentation and recognition of lesions can be time-consuming and inconsistent, with different clinicians potentially yielding

Table 5. The applications of U-Net++ in ultrasound image analysis

Presenters	Analytical objectives	Model	Results
Zhou et al. [41]	Carotid vessel wall	UNet++ ensemble approach	The Dice coefficients of the media-adventitia boundary (MAB) are 91.07%–91.56%, and those of the lumen-intima boundary (LIB) are 87.53%–89.44%.
Chen et al. [42]	Thyroid nodules	Based on the U-Net++ model, EfficientDet is used as the encoder, and the CSSE block is merged in the encoder and decoder.	Dice similarity coefficient of 0.9213
Orlando et al. [43]	Prostate	U-Net++	Dice similarity coefficient of 85%–89%
Zhu et al. [44]	lumen and median-adventitia	Feature pyramid network was added to the U-Net++ model.	Jaccard measure (JM) of 0.9080 ± 0.0321 and a Hausdorff distance (HD) of 0.1484 ± 0.1584 mm for the lumen border; JM of 0.9199 ± 0.0370 and HD of 0.1781 ± 0.1906 mm for the MA border.
Xu et al. [45]	Vessel	EAU-Net++: basis of U-Net++, EfficientNet is employed as an encoder.	The Accuracy of EAU-Net++ is 0.99.
Dai et al. [46]	Thyroid nodule	SK-Unet++: adds the adaptive receptive fields based on the selective kernel (SK) attention mechanisms into the Unet++ network.	Accuracy of 80.6%; Dice similarity coefficient of 84.7%
Zhou et al.[47]	Carotid plaques	UNet++ ensemble algorithm	The DSC in the SPARC dataset is 83.3%–85.7%; the DSC of the Zhongnan dataset is 88.6%.

Note: CSSE, Concurrent squeeze and spatial excitation; MA, Median Adventitia; DSC, Dice similarity coefficient.

varying results. Therefore, segmentation algorithms, like U-Net, have become crucial tools in standardizing and improving lesion detection [28]. A summary of U-Net's applications in ultrasound image analysis is presented in **Table 4**.

U-Net++

U-Net++, an advanced variant of the original U-Net, has demonstrated superior performance in various ultrasound applications. In obstetric ultrasound, the model excels in accurately outlining the morphology of fetal heart chambers and valves, as well as precisely delineating the placenta's boundaries. This significantly enhances the accuracy of fetal developmental assessments and the reliability of monitoring placental abnormalities. In abdominal ultrasound, U-Net++ effectively distinguishes tumors, cysts, and lesions in organs such as the liver and kidneys from normal tissues, providing a robust foundation for the development of subsequent treatment plans. Additionally, during ultrasound-guided puncture biopsy, U-Net++ can distinguish the target tissue from surrounding critical structures in real time, reducing procedure-related risks.

Furthermore, U-Net++ demonstrates high-precision segmentation in ultrasound images with blurred boundaries. For instance, in breast ultrasound tumor segmentation tasks, the model achieves a high Dice coefficient, enabling more

accurate tumor boundary delineation. Its multi-scale feature extraction capacity allows the model to capture both global information, such as the overall orientation of the carotid artery, and microscopic details, such as the fine structure of the intima, media, and plaques.

However, UNet++ also has limitations. The model requires substantial computational resources, and issues such as memory overflow may arise when processing large ultrasound image datasets on typical computer configurations, thereby limiting its applicability in resource-constrained environments. Additionally, the training process is also time-intensive, often taking days or even weeks to process large volumes of ultrasound images, significantly extending the research and development cycle and increasing costs. Furthermore, due to the nature of ultrasound images, U-Net++ may be susceptible to overfitting, leading to poor generalization performance when handling new or unseen images. A summary of U-Net++'s applications in ultrasound image analysis is presented in **Table 5**.

Attention U-Net

Compared to traditional U-Net models, Attention U-Net innovatively integrates an attention mechanism, which enhances region selectivity during the feature extraction. This innovation is particularly beneficial in ultrasound image analysis, such as thyroid nodule detection, where

Table 6. The applications of Attention U-Net in ultrasound image analysis

Presenters	Analytical objectives	Network	Results
Chen et al. [48]	Breast	Adaptive attention U-net	It has better generalization performance in the breast lesion segmentation.
Huang et al. [49]	Pleural effusion areas	Attention U-net	Average specificity of 91%, average Dice coefficient was 0.86.
Jain et al. [50]	Internal carotid artery and common carotid artery	Attention U-net	The correlation coefficient value f is 0.96; the AUC value is 0.97.
Katakis et al. [51]	Skeletal muscles	Attention U-net	Exhibits superb performance
Khaledyan et al. [52]	Breast	The novel approach by combining Sharp UNet and Attention UNet	The Dice coefficient, specificity, sensitivity, and F1 score values obtained were 0.93, 0.99, 0.94, and 0.94.
Shah et al. [53]	Brachial Plexus after clavicular surgeries	Attention U-net	The Dice coefficient improves about 1–2%.

Note: AUC, Area Under the Curve.

Table 7. The applications of ResU-Net in ultrasound image analysis

Presenters	Analytical objectives	Model	Results
Cai et al. [54]	Dipose layers	Attention Skip-Convolutions ResU-Net	It obtains a Dice of $99.06 \pm 0.95\%$ and an ASD of 0.19 ± 0.18 mm.
Dhar et al. [55]	Breast cancerous tissue	Connected ResU-Net (combine the algorithm of ResU-Net and Connected U-Net algorithms.)	IoU of 72.50% for the training and 63.10% for the testing dataset
Wang et al. [56]	Nerve of brachial plexus	ResU-Net	The dice coefficient 0.7093
Honghan et al. [57]	Breast tumor	Rms-se-U-Net	Dice similarity coefficient and intersection over union were 94.69% and 90.01%.

Note: ASD, Average Surface Distance.

the model can focus on key features like nodule boundaries, internal echogenic structures, and blood flow signals. For small papillary cancer nodules, Attention U-Net can accurately capture distinctive features such as gravel-like calcification and hypoechoic halos, significantly improving the accuracy of distinguishing between benign and malignant thyroid nodules and reducing the risk of misdiagnosis. Furthermore, Attention U-Net demonstrates strong resilience to noise and artifacts present in ultrasound images. In abdominal ultrasonography, noise introduced by factors like intestinal gas and patient movement can interfere with image clarity. The attention mechanism helps the model differentiate between true lesion features and noise artifacts, maintaining accurate lesion representation in organs such as the liver and pancreas, thus ensuring diagnostic accuracy.

Despite its advantages, Attention U-Net has certain limitations, particularly in its requirement for a large volume of accurately labeled ultrasound image data during training. Labeling ultrasound images is a challenging task that demands not only profound medical expertise but also a deep understanding of ultrasound imaging principles. Variations in labeling practices and standards among different personnel can lead to inconsistent data labeling quality, which may result in the model producing inac-

curate diagnostic results. A summary of Attention U-Net's application in ultrasound image analysis is presented in **Table 6**.

ResU-Net

ResU-Net is particularly effective in ultrasound applications that require high-precision segmentation, such as the automated segmentation of organs and lesions, including the liver, breast, heart, fetus, and kidneys. Ultrasound images often suffer from noise and artifacts, making traditional image segmentation methods struggle to accurately delineate the lesion regions, particularly in cases where tumors or organs exhibit fuzzy boundaries and minimal intensity differences. In such instances, ResU-Net leverages its residual connections to improve feature flow and gradient propagation, thereby enhancing the model's capacity to learn fine details. This results in more accurate segmentation, even in complex or noisy backgrounds with blurred boundaries. For instance, in the segmentation of liver ultrasound images, the boundaries between the liver and surrounding tissues are often unclear and prone to noise. ResU-Net effectively separates the liver contours from these blurred regions, supporting more accurate diagnosis of liver diseases, such as hepatocellular carcinoma and fatty liver. In breast ultrasound images, where tumor bound-

aries are similarly indistinct, ResU-Net, by learning residual features, not only achieves precise tumor segmentation but also demonstrates high sensitivity in detecting small lesions or early-stage tumors. In fetal ultrasound images, accurate segmentation of fetal organs is crucial. ResU-Net's ability to extract fine-grained features allows clinicians to better visualize critical organs such as the fetal heart and brain, providing valuable diagnostic information.

However, the application of ResU-Net in ultrasound images also encounters challenges. For example, when dealing with low-quality, low-resolution, or severely noise-contaminated ultrasound images, despite enhanced gradient flow, the model may still struggle with inaccurate segmentation, missed detections, or excessive artifacts when dealing with extreme blurriness or insufficient training samples. In such cases, combining ResU-Net with other image enhancement techniques or multimodal data fusion could improve its robustness and segmentation performance. A summary of ResU-Net's application in ultrasound image analysis is presented in **Table 7**.

III Conclusion

This review summarizes the application of U-Net-based models in ultrasound imaging, highlighting their significant contributions to medical diagnosis. Variants such as U-Net++, Attention U-Net, and ResU-Net have significantly enhanced both accuracy and efficiency in image segmentation tasks. Continuous advancements in segmentation-related research have been made, including the utilization of methods such as multi-scale feature pyramids and multi-level attention pyramids [58]. These approaches provide fine-grained feature extraction by focusing on local details and improving the representation of critical regions, thus offering valuable insights for enhancing segmentation performance.

Despite these advancements, challenges remain in ensuring the generalization and robustness of U-Net models across diverse datasets and varying imaging conditions. Future research should prioritize enhancing the adaptability of U-Net models to various ultrasound imaging environments while improving their interpretability for clinical use. Additionally, the integration of U-Net models with advanced techniques like transfer learning and multimodal imaging is anticipated to further expand their applicability and performance.

Author contributions: Yuxiang Wang: Respon-

sible for the general framework design of the article, the planning and organization of the research direction, led the writing and editing of the literature review section, and participated in the discussion and writing of the research methodology section. Miao Zhou: Provided medically relevant theoretical support and was responsible for the collection and organization of literature. Fangfang Chen: Participated in literature searches, provided experimental data support, and was responsible for producing some of the graphs. Jintao Duan: Partial data collection and organization and production of partial tables. Liangqing LIN: Provides theoretical support for relevant parts of the development of ultrasound medicine. Qinghua Wu: Language optimization and restructuring of the sections of the article. Wenhui Guo: The analysis of the research methodology and technical details is discussed in depth and the structure of the article is optimized. Haipo Cui: Led the revision of the article, participated in the final review and revision of the article, ensured that all authors' comments were integrated, and ultimately approved the article for submission.

References

- [1] Shin SJ, Jeong BJ. Principle and comprehension of ultrasound imaging. *J Korean Orthop Assoc* 2013;48(5):325-333.
- [2] Pehrson LM, Lauridsen C, Nielsen MB. Machine learning and deep learning applied in ultrasound. *Ultraschall Med* 2018;39(4):379-381.
- [3] Micucci M, Iula A. Recent advances in machine learning applied to ultrasound imaging. *Electronics* 2022;11(11):1800.
- [4] Yap MH, Pons G, Marti J, et al. Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE J Biomed Health Inform* 2017;22(4):1218-1226.
- [5] Almajalid R, Shan J, Du Y, et al. Development of a deep-learning-based method for breast ultrasound image segmentation. 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA) 2018;1103-1108.
- [6] Shareef B, Vakanski A, Freer PE, et al. Estan: Enhanced small tumor-aware network for breast ultrasound image segmentation. *Healthcare (Basel)*. 2022;10(11):2262.
- [7] Hossain S, Azam S, Montaha S, et al. Automated breast tumor ultrasound image segmentation with hybrid UNet and classification using fine-tuned CNN model. *Heliyon* 2023;9(11):e21369.
- [8] Weng Y, Zhou T, Li Y, et al. Nas-unet: Neural architecture search for medical image segmentation. *IEEE Access* 2019;7:44247-

- 44257.
- [9] Zhong LF, Han M. A Unet-based method for Cell Segmentation Challenge. 36th Conference on Neural Information Processing Systems (NeurIPS 2022). 2022;1-5.
- [10] Zeng Z, Xie W, Zhang Y, et al. RIC-Unet: An improved neural network based on Unet for nuclei segmentation in histology images. IEEE Access 2019;7:21420-21428.
- [11] Tang P, Liang Q, Yan X, et al. Efficient skin lesion segmentation using separable-Unet with stochastic weight averaging. Comput Methods Programs Biomed 2019;178:289-301.
- [12] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18 2015;234-241.
- [13] Dubey SR, Singh SK, Chaudhuri BB. Activation functions in deep learning: A comprehensive survey and benchmark. Neurocomputing 2022;503:92-108.
- [14] Sharma S, Sharma S, Athaiya A. ACTIVATION FUNCTIONS IN NEURAL NETWORKS. Int J Eng Appl Sci Technol 2020;12(4):310-316.
- [15] Dumoulin V, Visin F. A guide to convolution arithmetic for deep learning. arXiv preprint arXiv:1603.07285 2016.
- [16] Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2 2014;2672–2680.
- [17] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2015;3431-3440.
- [18] Noh H, Hong S, Han B. Learning Deconvolution Network for Semantic Segmentation. 2015 IEEE International Conference on Computer Vision (ICCV) 2015;1520-1528.
- [19] Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, et al. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support 2018;3-11.
- [20] Lee CY, Xie S, Gallagher P, et al. Deeply-Supervised Nets. Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics 2015;562-570.
- [21] Zhou Z, Siddiquee MMR, Tajbakhsh N, et al. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE Trans Med Imaging 2019;39(6):1856-1867.
- [22] Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning. Neurocomputing 2021;452:48-62.
- [23] Guo MH, Xu TX, Liu JJ, et al. Attention mechanisms in computer vision: A survey. Comput Visual Media 2022;8(3):331-368.
- [24] Oktay O, Schlemper J, Folgoc LL, et al. Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 2018.
- [25] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition 2016;770-778.
- [26] Xiao X, Lian S, Luo Z, et al. Weighted res-unet for high-quality retina vessel segmentation. 2018 9th international conference on information technology in medicine and education (ITME) 2018;327-331.
- [27] International Agency for Research on Cancer. (2012). GLOBOCAN 2012: Estimated cancer incidence, mortality and prevalence worldwide in 2012. Available at: <https://gco.iarc.fr/en>
- [28] Li Y, Chouzenoux E, Charmettant B, et al. Lightweight u-net for lesion segmentation in ultrasound images. 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI) 2021;611-615.
- [29] Zyuzin V, Sergey P, Mukhtarov A, et al. Identification of the left ventricle endocardial border on two-dimensional ultrasound images using the convolutional neural network Unet. 2018 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBREIT) 2018;76-78.
- [30] Wang Y, Geng J, Zhou C, et al. Segmentation of ultrasound brachial plexus based on U-Net. 2021 international conference on communications, information system and computer engineering (CISCE) 2021;482-485.
- [31] Tian D, Zhu B, Wang J, et al. Brachial plexus nerve trunk recognition from ultrasound images: a comparative study of deep learning models. IEEE Access 2022;10:82003-82014.
- [32] Li X, Hong Y, Kong D, et al. Automatic segmentation of levator hiatus from ultrasound images using U-net with dense connections. Phys Med Biol 2019;64(7):075015.
- [33] Díaz-Vargas HM, Jimenez-Castaño CA, Cárdenas-Peña DA, et al. Peripheral Nerve Segmentation in Ultrasound Images Using Conditioned U-Net. Prog Artif Intell Pattern Recognit 2021;124-133.
- [34] Zou Y, Amidi E, Luo H, et al. Ultrasound-enhanced Unet model for quantitative photoacoustic tomography of ovarian lesions. Photoacoustics 2022;28:100420.
- [35] Yang J, Faraji M, Basu A. Robust segmentation

- of arterial walls in intravascular ultrasound images using Dual Path U-Net. *Ultrasonics* 2019;96:24-33.
- [36] Ben Ahmed I, Ouarda W, Ben Amar C. Hybrid UNET model segmentation for an early breast cancer detection using ultrasound images. *International Conference on Computational Collective Intelligence* 2022;464-476.
- [37] Shodiq MN, Yuniarno EM, Nugroho J, et al. Ultrasound image segmentation for deep vein thrombosis using UNet-CNN based on denoising filter. *2022 IEEE international conference on imaging systems and techniques (IST)* 2022;1-6.
- [38] Suissa N. Development and validation of automatic ultrasound nerve block guidance for regional anesthesia using deep learning. McGill University 2023.
- [39] Oghli MG, Moradi S, Sirjani N, et al. Automatic measurement of fetal head biometry from ultrasound images using deep neural networks. *2020 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)* 2020;1-3.
- [40] Lin S, Zhao Y, Lei D, et al. Perioperative Nursing Management of Patients Undergoing Laparoscopic Ovarian Cystectomy Guided by Ultrasound Imaging under Intelligent Algorithm. *Comput Math Methods Med* 2022;2022:7193005.
- [41] Zhou R, Guo F, Azarpazhooh MR, et al. Carotid vessel-wall-volume ultrasound measurement via a UNet++ ensemble algorithm trained on small data sets. *Ultrasound Med Biol* 2023;49(4):1031-1036.
- [42] Chen C, Xu B, Wu Y, et al. Research on ultrasonic image segmentation of thyroid nodules based on improved U-net++. *2022 2nd International Conference on Bioinformatics and Intelligent Computing* 2022;532-536.
- [43] Orlando N, Gyacskov I, Gillies DJ, et al. Effect of dataset size, image quality, and image type on deep learning-based automatic prostate segmentation in 3D ultrasound. *Phys Med Biol* 2022;67(7):074002.
- [44] Zhu F, Gao Z, Zhao C, et al. A deep learning-based method to extract lumen and media-adventitia in intravascular ultrasound images. *Ultrason Imaging* 2022;44(5-6):191-203.
- [45] Xu X, Luan K, Liu J. EAU-Net++: An Efficient Segmentation Network for Ultrasound Vessel Images. *2023 IEEE International Conference on Mechatronics and Automation (ICMA)* 2023;1809-1814.
- [46] Dai H, Xie W, Xia E. SK-Unet++: An improved Unet++ network with adaptive receptive fields for automatic segmentation of ultrasound thyroid nodule images. *Med Phys* 2024;51(3):1798-1811.
- [47] Zhou R, Guo F, Azarpazhooh MR, et al. Deep learning-based measurement of total plaque area in B-mode ultrasound images. *IEEE J Biomed Health Inform* 2021;25(8):2967-2977.
- [48] Chen G, Li L, Dai Y, et al. AAU-net: an adaptive attention U-net for breast lesions segmentation in ultrasound images. *IEEE Trans Med Imaging* 2022;42(5):1289-1300.
- [49] Huang L, Lin Y, Cao P, et al. Automated detection and segmentation of pleural effusion on ultrasound images using an Attention U-net. *J Appl Clin Med Phys* 2024;25(1):e14231.
- [50] Jain PK, Dubey A, Saba L, et al. Attention-based UNet deep learning model for plaque segmentation in carotid ultrasound for stroke risk stratification: an artificial intelligence paradigm. *J Cardiovasc Dev Dis* 2022;9(10):326.
- [51] Katakis S, Barotsis N, Kakotaritis A, et al. Automatic extraction of muscle parameters with attention UNet in ultrasonography. *Sensors* 2022;22(14):5230.
- [52] Khaledyan D, Marini TJ, M. Baran T, et al. Enhancing breast ultrasound segmentation through fine-tuning and optimization techniques: sharp attention UNet. *Plos one* 2023;18(12):e0289195.
- [53] Shah N, Bhatia J, Vasavat N, et al. Ultrasound Nerve Image Segmentation Using Attention Mechanism. *International Conference on Advances and Applications of Artificial Intelligence and Machine Learning* 2022;789-802.
- [54] Cai R, Liu Y, Sun Z, et al. Deep-learning based segmentation of ultrasound adipose image for liposuction. *Int J Med Robot* 2023;19(6):e2548.
- [55] Dhar O, Yow K-C. Connected ResU-Net: A Deep Learning Model for Segmentation of Breast Cancer Ultrasound Images. *2023 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)* 2023;88-93.
- [56] Wang R, Shen H, Zhou M. Ultrasound nerve segmentation of brachial plexus based on optimized ResU-Net. *2019 IEEE international conference on imaging systems and techniques (IST)* 2019;1-6.
- [57] Honghan Z, Liu DC, Jingyan L, et al. Rms-se-unet: A segmentation method for tumors in breast ultrasound images. *2021 IEEE 6th International Conference on Computer and Communication Systems (ICCCS)* 2021;328-334.
- [58] Tang H, Yuan C, Li Z, et al. Learning attention-guided pyramidal features for few-shot fine-grained recognition. *Pattern Recognit* 2022;130:108792.