TECHNICAL SCIENCES

AUTOMATED SEGMENTATION OF ULTRASOUND MEDICAL IMAGES USING THE ATTENTION U-NET MODEL

Momot A.,
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute, PhD, Senior Lecturer
Zaboluieva M.,
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute, Student
Galagan R.
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute, PhD, Assistant Professor
https://doi.org/10.5281/zenodo.10817342

Abstract

The article deals with the method of automated semantic segmentation of ultrasound medical images using the Attention U-Net deep learning model. The advantages of using Attention blocks in neural network architectures for segmentation tasks are analyzed. To test the described algorithms, the Breast Ultrasound Images training dataset was chosen. The method described in the article allows for automating the process of detecting and preliminary classification of breast tumors based on the analysis of ultrasound images. As a result of training the Attention U-Net model, the Mean IOU value of 49.2% was obtained on the test set. The network can automatically classify the detected neoplasm as benign or malignant with an F1 Score of 0.87. The results indicate the prospects of using the Attention U-Net model in the tasks of analyzing ultrasound medical images. Ways to further improve the considered method are proposed.

Keywords: deep learning, segmentation, ultrasound, medical diagnostics, breast cancer.

Introduction

Today, ultrasound diagnostics (US) is an essential tool in almost all areas of medicine. Ultrasound diagnostics (ultrasound) is a procedure that involves the application of high-frequency sound waves to a body part and subsequent imaging of internal organs. Ultrasound diagnostics, unlike X-ray diagnostics, does not use dangerous ionizing radiation. Since the image obtained by ultrasound is displayed in real-time, such diagnostics allows us to determine the structure and analyze the movement of internal organs and blood entering the blood vessels [1].

Image segmentation is the process of dividing an image into its parts or objects in the image, i.e. sets of pixels [2]. To find and identify the boundaries of objects in an image, pixels are evaluated according to some homogeneous criteria (color, intensity or texture).

The accelerated development in medical image processing is mainly due to the use of deep learning technologies [3], which allow for effective learning of object features directly from imaging data. In medical diagnostics, neural networks make it possible to significantly improve the accuracy of diagnosis and spend less time on analysis. Services based on neural networks are being developed all over the world to help doctors detect various pathologies and diseases, including oncology.

Image segmentation based on machine learning technologies is already firmly established as a reliable tool for medical image analysis. However, automated segmentation of medical images is a challenging task due to several problems, such as a large variety of anatomy shapes and sizes between patients, low contrast with surrounding tissues, and the lack of a large amount of digital medical data in the public domain. The objective of this paper is to analyze the effectiveness of the Attention U-Net model for automated detection and extraction of anomalies in ultrasound (US) images.

Review of recent works

The semantic segmentation algorithm is based on the fact that each pixel in the image is assigned to a certain class depending on what object it belongs to. This creates masks, which are parts of the image differentiated from a specific area. The main models used in conjunction with this segmentation method are deep neural networks, namely fully convolutional neural networks. Since the process of increasing the network layer leads to a decrease in spatial information, a special sampling layer was created, which makes it possible to optimize the image by reducing or increasing the sampling (downsampling, upsampling) [4]. Another important innovation was the max-pooling tool, which is responsible for extracting information from image areas and analyzing them.

Despite the power of this algorithm and its many advantages, it also has some disadvantages that affect the use of this method in medical image segmentation tasks. Classical deep neural networks used for semantic segmentation have low edge resolution due to information loss during the encoding process. The use of convolutional neural networks requires large computing resources, powerful hardware, and a significant amount of data for training, which often becomes an obstacle to using them in some tasks. Nevertheless, the application of these deep algorithms has great potential.

In medical diagnostics, neural networks can significantly improve the accuracy of diagnosis and spend less time on analysis.

The authors of [5] presented an algorithm for classifying ultrasound images. VGG16 was chosen as the architecture. Having analyzed the results of network training, we can say that the accuracy of the model on the training data is high, but on the validation data it is low. This may indicate overtraining. Even though VGG16 has many advantages and shows good results in image classification tasks, it has too low a learning rate and too much weight. In addition, this architecture is not designed for object detection or segmentation. Due to these shortcomings, it is important to use more advanced convolutional neural network architectures.

In [6], classification and detection algorithms based on convolutional neural networks trained on ultrasound images were developed and analyzed. In this experiment, the authors used the following architectures: AlexNet, ZFNet, VGG16, GoogLeNet, ResNet, and DenseNet. As a result of training, the network with the DenseNet architecture showed a correct answer rate of 80.0% on the test data. This level of reliability cannot be acceptable for making a final diagnosis. Therefore, the proposed model needs to be improved. In addition, this work solved only detection and classification tasks. This is ineffective in the field of ultrasound image analysis, as it does not allow for an accurate assessment of the area of pathological areas.

Paper [7] considers the application of semantic segmentation of ultrasound images using deep neural networks. The authors used the DeepLab v3+ architecture based on ResNet-50. According to the results obtained, the reliability of the system on the test data was 97.58%, which is a fairly good result. However, the training graph shows that the system was overtrained. The IoU (Intersection over Union) metric for assessing the quality of segmentation has a value of 81.12%. The disadvantage of this model is its complexity due to the large number of parameters.

The authors of [8] consider an algorithm for segmenting ultrasound images using the U-Net architecture. The network has an architecture resembling the letter "U", with an encoder on the left side and a decoder on the right. The encoder is responsible for the stepwise compression and extraction of abstract image features, while the decoder is responsible for the stepwise recovery of the spatial size and details of the segmentation map. The reliability of the system was very high, at 99.05%, but of course, false positives were also observed.

Despite the good result, this neural network model was applied only to small data sets, which is why the efficiency of this system was very low. At the moment, there are many innovative architectures based on the basic U-Net, retaining all the advantages of this algorithm, and at the same time, with improved characteristics and eliminated disadvantages. The right decision to improve this system would be to change the architecture towards a more promising one for medical image segmentation tasks.

Attention U-Net model

One of the best models successfully applied in medical image segmentation tasks is the U-Net architecture network. The U-Net architecture consists of two parts: a narrowing part (encoder) and an expanding part (decoder). The encoder gradually reduces the spatial dimension by merging layers, and the decoder gradually restores object details and spatial dimension. In the narrowing (input) part of the structure, the incoming image passes through a series of layers: convolution and subsampling (pooling) layers [9].

The network does not have fully connected layers and uses only the valid part of each convolution, i.e., the segmentation map contains only pixels for which the full context is available in the input image. For high-quality segmentation, U-Net increases the amount of data by deforming existing images using the MaxPooling operation, the main idea of which is to reduce the number of parameters and computations and present images that are more resistant to displacement and scaling, while preserving all their features.

The authors of [10] developed algorithms for the segmentation of ultrasound images of the breast by various neural networks: FCN, Mask R-CNN, and U-Net. Comparing the results, it can be concluded that the U-Net neural network performed the task with the highest quality. It accurately conveyed the shape and size of objects. As a result of the training, it showed a correct answer rate of 84.6% on the test data, which was the best result among the models under consideration.

Today, there are many new models based on the basic U-Net that have better performance and are more efficient in image segmentation tasks. We propose to consider a modification of this model - Attention U-Net. Recently, this model has been showing the best results in segmentation tasks.

"Attention", in the context of image segmentation, is a way to extract only relevant activations during training. This reduces the number of computational resources wasted on irrelevant activations, providing the network with better computational efficiency. In other words, the network can pay "attention" to certain details of the image that are most important for decision-making [11]. The structure of the Attention Gate block is shown in Figure 1.

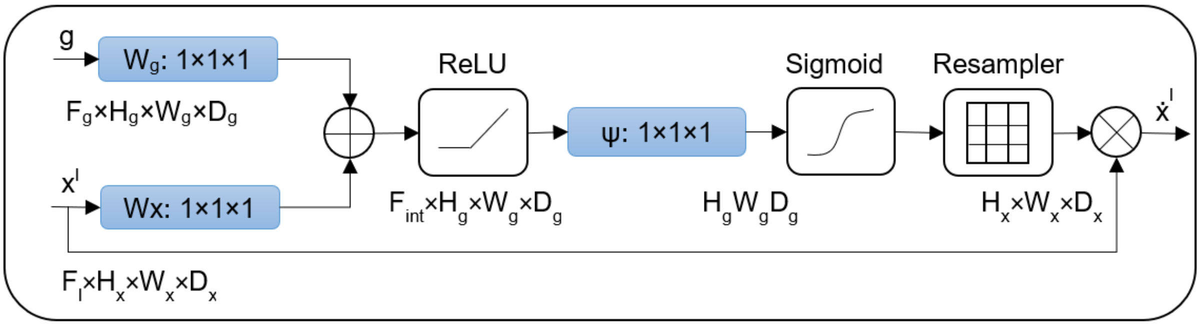


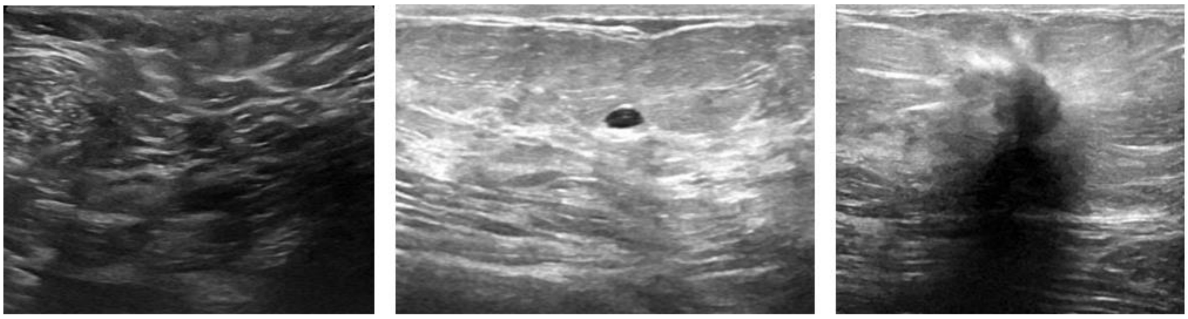
Figure 1. Structure of the Attention Gate block

Attention can come in two forms: hard and soft. The hard form works based on selecting relevant areas by framing the image or iteratively proposing an area. Since hard form can only select one area of an image at a time, two disadvantages arise: it is not differentiated and requires reinforcement learning to train. Soft form works by weighting different parts of the image. Areas of high relevance are multiplied by a higher weight, and areas of low relevance are tagged with lower weights. When the model is trained, more attention is paid to the areas with higher weights. Unlike a rigid shape, these weights can be applied to many areas of an image.

The Attention U-Net model has improved segmentation accuracy. The results obtained by the authors of [12] show that Attention U-Net outperforms the baseline U-Net in terms of the overall Dice coefficient by a significant margin. And, despite having more parameters, it has the same learning speed. The Attention block makes it possible to pay special attention to certain areas of interest, work with complex images, and perform control at different depth levels. These are the aspects that make the Attention U-Net effective for medical images.

Results and discussion

To train the neural network, we propose to use the open dataset "Breast ultrasound images" [13]. Examples of images from the training dataset are shown in Figure 2.



a b c

Figure 2. Examples of ultrasound images of the breast structure:
a - no pathology, b - benign tumor, c - malignant tumor

The dataset contains 780 ultrasound images obtained during the diagnosis of breast pathologies in 600 women aged 25 to 75 years. The images are divided into three classes: no pathology, benign and malignant tumors. In the future, by increasing the dataset with new ultrasound images, the system's reliability will be improved.

When developing segmentation models, it is important to choose the right metrics to evaluate their effectiveness. Let's analyze a few key metrics that are commonly used.

Accuracy measures the overall performance of the model by comparing the number of correctly classified pixels to the total number of pixels. Accuracy is a good indicator for an overall overview of the model's performance but may be unreliable in the case of unbalanced data [14].

F1 Score is a metric that is used when classes are unbalanced. It is used in cases where you need to find a balance between accuracy and completeness. In image segmentation tasks, the F1 Score is a more reliable indicator than accuracy. This is because a characteristic feature of medical image segmentation is the dominance of the background class. The background pixels do not belong to any objects, but they affect the metrics.

The IoU (Intersection over Union) metric is characterized by the proportion of overlap between the predicted segmentation area and the actual segmentation area before the two areas are merged. IoU is a key tool for segmentation tasks because it illustrates how accurately a model can identify and separate different objects. Currently, there are many modifications of this metric, including Mean IoU, the main difference of which is that it calculates the average IoU value over the entire data set [15]. This metric is the most suitable for the Attention U-Net architecture.

The Transfer learning approach was used to accelerate learning. The Attention U-Net network used in the system was pre-trained on the ImageNet dataset. The architecture of the model consists of a decoder block, an encoder block, and Attention Gate blocks. The metric chosen is Accuracy, the optimizer is Adam, and the loss function is binary crossentropy. The network training results are presented in Table 1. The table includes the reduction of the loss function and changes in the accuracy metrics during training.

Table 1.

Learning results

Epoch	Loss	Accuracy	F1 Score	Mean IoU
1	0.40	0.90	0.61	0.455
5	0.20	0.92	0.72	0.455
10	0.17	0.93	0.78	0.457
15	0.15	0.95	0.83	0.491
20	0.13	0.97	0.87	0.492

The model showed a steady improvement over all 20 epochs, which indicates the effectiveness of the chosen parameters and architecture. The greatest improvement was observed in the first 15 epochs.

The Mean IOU metric showed a value of 49.2%. As a result of training, the network showed a 97% correct answer rate on the test data. The F1 Score was 0.87. Despite the overall improvement, the Mean IoU remained relatively low, which may indicate the need for further optimization of the model or training data set.

As a result of training, the system collectively showed the majority of true results. In Figure 3, you can see the tumor, which is colored yellow and correctly classified as a malignant tumor. The network has clearly drawn the contour shape and dimensions that match the test ones.

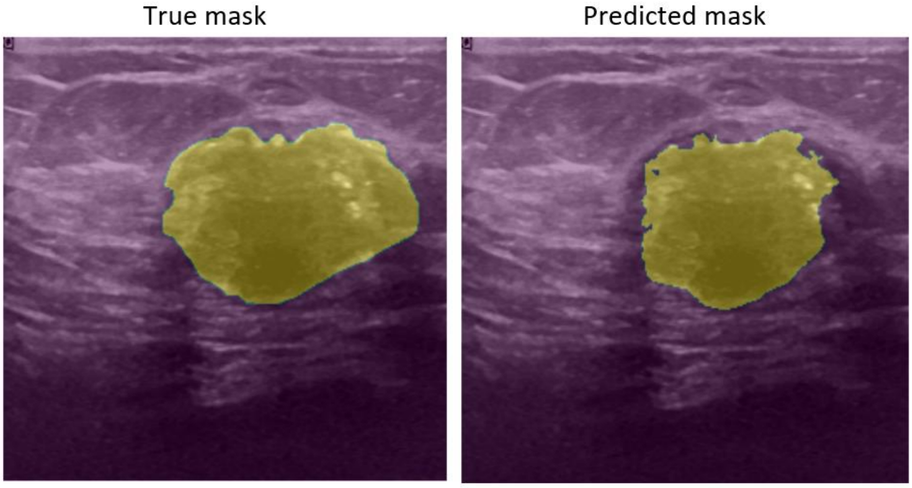


Figure 3. Results of ultrasound image segmentation of a malignant tumor

There are also cases of false results. In Figure 4, the system has incorrectly identified an object that is a malignant neoplasm and has a complex branched shape. This confirms that the system is not perfect and needs to be improved.

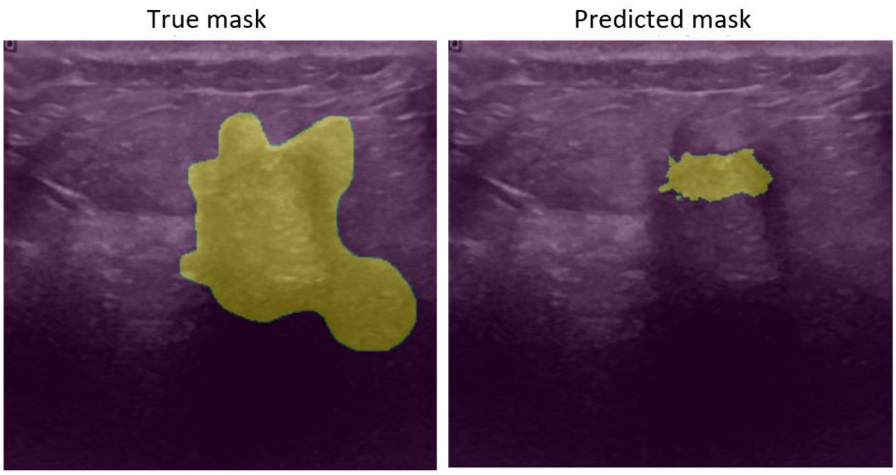


Figure 4. Results of inaccurate segmentation of a malignant tumor

Despite the presence of false positives, most of them are correct and the system showed a high proportion of correct answers. The use of deep learning models has a number of advantages over classical segmentation methods, but there are also limitations. Among the existing limitations is that a large number of ultrasound images are required to train a neural network. An insufficient number of images or their low representativeness leads to a deterioration in the quality of segmentation, which is a common drawback for all deep learning models. Thus, an important task is to expand the training data set. In particular, increasing the resolution of images, increasing the number of object classes, and the quality of their annotations.

Conclusion

Ultrasound medical images are challenging objects for automated analysis. As a rule, they contain high levels of interference, the contours of objects in the images are fuzzy, and objects can have a complex branched structure and overlap. Therefore, to automate the process of processing such images, it is promising to use artificial intelligence methods. In particular, deep learning.

Among the existing deep learning models for solving image segmentation problems, various modifications of U-Net are the most popular. The Attention U-Net architecture can be distinguished from the latest such modifications. Due to the use of Attention blocks, this network has increased efficiency in segmentation tasks. Therefore, the study of the capabilities of this model for the segmentation of ultrasound medical images is of considerable scientific interest.

This study analyzed the qualitative and quantitative results of the Attention U-Net model in the task of segmenting ultrasound images of breast pathologies. As a result of training, the network showed a 97% correct answer rate on the test data. The F1 Score of 0.87 and Mean IOU of 49.2% were achieved. Compared to the results of medical image segmentation obtained in similar studies, these indicators are not the best. This can be explained by the imperfection of the training dataset and the complexity of the task. It is also worth noting that no image preprocessing was performed in this study.

According to the predicted masks of neoplasms, it can be said that the system generally shows true results and the considered segmentation method can be used as a decision support system in medical diagnostics. We also identified promising areas for improving the system, including expanding the amount of training data to increase the adaptability of the trained model. In addition, future research could be aimed at improving the segmentation accuracy by modifying the architecture of the deep learning model.

References:

1. Theoretical foundations of ultrasonic non-destructive testing / R. M. Galagan. - Kyiv: Igor Sikorsky Kyiv Polytechnic Institute, 2019. - 263 p. (in Ukrainian).
2. Steshenko Y. V. Automation of the process of segmentation of images of metal surface defects using the neural network U-Net / Y.V. Steshenko, A.S. Momot, A.G. Protasov O.V. Muraviov
3. Ultrasound examination of the mammary glands. Step by step. From simple to complex: textbook / A.N. Sencha. - Kyiv: MEDpress-inform, 2018. 184 p.
4. Steger C. Machine vision algorithms and applications / Carsten Steger, Markus Ulrich, Christian Wiedemann: Wiley & Sons, Incorporated, John, 2017. - 516 p.
5. Zhou, Xiangrong, et al. // Deep learning of the sectional appearances of 3D CT images for anatomical structure segmentation based on an FCN voting method // Medical physics, - 2017. - pp. 5221-5233.
6. Long J. Fully convolutional networks for semantic segmentation [Electronic resource] / Jonathan Long, Evan Shelhamer, Trevor Darrell // 2015 IEEE conference on computer vision and pattern recognition (CVPR), Boston, MA, USA, 7-12 June 2015.
7. Semantic Segmentation with Extended DeepLabv3 Architecture [Electronic resource] // IEEE Xplore.
8. Ronneberger, Olaf, Philipp Fischer, and Thomas Brox: Convolutional networks for biomedical image segmentation // International Conference on Medical image computing and computer-assisted intervention. - Springer, Cham, - 2015. - pp. 155-170.
9. Manjunath, R. V.; Kwadiki, Karibasappa. // Modified U-NET on CT images for automatic segmentation of liver and its tumor. Biomedical Engineering Advances, - 2022. - pp. 103-110.
10. Ronneberger, Olaf, Philipp Fischer, and Thomas Brox: Convolutional networks for biomedical image segmentation // International Conference on Medical image computing and computer-assisted intervention. - Springer, Cham, - 2015. - pp. 155-170.
11. Attention U-Net Based Adversarial Architectures for Chest X-ray Lung Segmentation [Electronic resource] // arXiv.org.
12. Oktay, Ozan, et al. Attention u-net: Learning where to look for the pancreas. // arXiv preprint. - 2018. - pp. 99-119.
13. Al-Dhabyani W, Gomaa M, Khaled H, Fahmy A. A dataset of breast ultrasound images. Data in Brief. 2020 Feb; 28:104863.
14. Momot A. Deep learning automated data analysis of security infrared cameras / A. Momot, I. Skladchykov // Slovak international scientific journal. - 2021. - No. 52. - pp. 13-16.
15. Momot A. Analysis of the influence of the threshold level of binarisation on the efficiency of segmentation of images of surface defects in steel by the U-Net network / A. Momot, Y. Steshenko // XXII International Scientific and Technical Conference "Instrumentation: State and Prospects", 16-17 May 2023, Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine: conference proceedings.