Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 11, No. 2, 2025

Feature Recognition of Ultrasound Breast Images Based on Improved DSA-U++

Weichun Guan, Yuanyuan Zhang *, Xinshu Lv, Shuyun Lv, Mao Lin, Zhengyang Liu

University of Science and Technology Liaoning, Anshan Liaoning, 114001 China

*** Corresponding author:** Yuanyuan Zhang

Abstract: The aim of this project is to recognize features in ultrasound breast images. And an improved DSA-U++ model is proposed based on the image classification task, the traditional U-Net++ in the encoder part in the face of ultrasound images there is a lack of feature extraction, to solve this problem we use Resnet50 as the backbone of the extraction, in order to further enhance the ability of the feature learning, we also introduced the ASPP module, to help capture contextual information at different scales, and a module R-AS is designed to enhance the model multi-scale perception ability. information to enhance the model's multi-scale perception ability and a module R-AS. the output of R-AS after five stages is used as the encoding part of U-Net++, and the feature information extracted by the encoder is reconstructed and enhanced in the decoder part. In order to reduce the computational complexity and the number of parameters, and to maintain a certain feature extraction ability, we replace the traditional convolution in the decoder part of U-Net++ with a depth-separable convolution, which is experimentally validated on the AI Algorithm Elite Challenge dataset, which consists of ultrasound breast images with four features, namely, orientation, edges, calcifications, and shapes, and is trained with the DSA-U++ model. After the DSA-U++ model is trained, it improves on several indexes, not only improving the recognition ability of subtle features in ultrasound breast images, but also effectively improving the performance of image classification.

Keywords: Ultrasound Breast Imaging; Resnet; Depth-separable Convolution; U-Net++.

1. Introduction:

In 2021, breast cancer has surpassed lung cancer as the most frequently diagnosed cancer worldwide, and globally, especially in women early detection and accurate diagnosis are important to improve patient survival [1]. And the World Health Organization's International Agency for Research on Cancer has released shows that it is expected to be increasing by 2070 [2, 3]. Among the many imaging methods, ultrasound imaging has become an important tool for the detection of breast diseases due to its noninvasiveness, real-time nature, and good discriminatory ability for dense breast tissue. However, due to the inherent limitations of ultrasound images themselves and the complexity of breast anatomy, its application in breast tumor detection and diagnosis still faces many challenges.

First, the quality of ultrasound images and the complexity of anatomical structures greatly affect the accuracy of detection. The quality of ultrasound images is constrained by a variety of factors, such as the depth of acoustic wave penetration, the density of the breast, and the scattering of the acoustic beam. Breast tissue consists of a variety of components such as fat, glands, and fibers, and its complex anatomical structure is prone to form shadows, artifacts, and unclear structures in the image, which makes accurate detection and localization of tumors extraordinarily difficult. Secondly, surrounding tissues, shape, margin contour, lesion boundaries, and posterior acoustic features are important factors to be considered in classifying a lesion [4]. The morphology of breast tumors may vary from individual to individual; for example, some tumors are extremely small and irregularly shaped, while others may be highly similar to the surrounding tissue. The accuracy for identifying tumors of different sizes also varies [5], and this diversity not only increases the difficulty of tumor detection in ultrasound images, but also puts higher demands on the robustness of the model. In addition, the similarity of benign and malignant breast tumor characteristics further exacerbates the diagnostic complexity. Certain benign and malignant tumors may present similar morphological features in ultrasound images, such as clear boundaries or homogeneous internal echoes, etc., which poses a great challenge to accurately differentiate benign and malignant tumors. Finally, noise and artifacts in ultrasound images are not to be ignored. False-positive or false-negative results often appear in ultrasound images due to interfering factors such as instrument noise, artifacts, and patient motion artifacts, leading to misdetections or missed detections.

Deep learning can learn more image features from a large amount of image data, which is of great significance in the field of image processing [6]. It has been widely used in fields such as computer vision [7, 8] and natural language processing [9, 10]. However, traditional ultrasound image diagnosis based on texture and morphological features of its images, which are manually analyzed by experienced doctors [11], shows some limitations in the recognition of breast lesions. With the rapid development of deep learning technology, many remarkable results have been achieved in the research of medical image processing [12, 13].

U-Net++, as a classical convolutional neural network architecture for image segmentation tasks, significantly enhances the network's ability in multi-scale feature fusion and detail recovery by introducing nested bar hopping connections and dense hopping connection structure on the basis of traditional U-Net. It is widely used due to its finer structure and excellent segmentation performance. In this project, we apply it to an image classification task to capture richer and more diverse features using its dense connections. However, U-Net++ is in the encoder part, and its feature extraction capability is insufficient when facing complex and

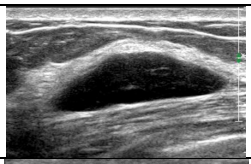
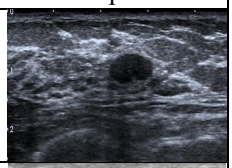
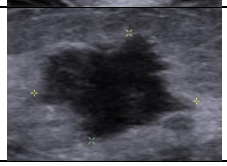
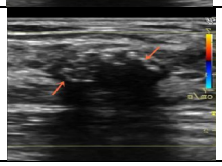
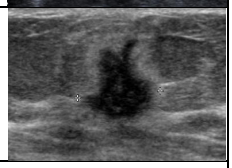
detail-rich images. We design an R-AS encoder module that employs Resnet50 as its backbone network, which utilizes its deep residual structure to enhance the feature extraction capability while effectively solving the problem of gradient vanishing. In addition, we also introduce a null-space pyramid pooling module for enhancing the model's ability to perceive features at different scales. The high computational complexity of the structure of U-Net++ leads to slow inference, especially when applied on resource-limited devices. In order to reduce the computational complexity and improve the inference efficiency, we replace some of the traditional convolutions in U-Net++ with depth-separable convolutions.

2. Related Work:

2.1. Data Sources:

The dataset in this paper is derived from the AI Algorithm Elite Challenge competition dataset, which is labeled with the assistance of professional physicians to ensure the accuracy of data labeling. The breast feature dataset contains four types of features: orientation, edge, calcification and shape, and each type of feature is represented by a binary label (0 and 1), where 0 indicates benign features and 1 indicates malignant features. As shown in Table 1.

Table 1. Demonstration of breast characterization dataset

	direction	**boundary**	**calcification**	shape
benign				
malignant				
descriptive	**Parallel: The long axis of the mass is parallel to the skin, i.e. horizontal position. Non-parallel: the anterior-posterior diameter of the mass is larger than the transverse diameter, i.e., vertical position.**	**Luminescence: the mass has well-defined borders and there is a distinct mutation between the lesion and the surrounding tissue. Unglossy: the mass has blurred borders and may exhibit a hyperechoic halo, angularity, microfoliation, or burr-like appearance.**	**Calcification: localized areas of highlighting (calcified spots) are visible in the ultrasound image No calcification: no visible areas of highlighting in the ultrasound image.**	**Regular: the mass is round, oval or large lobulated. Irregular: the shape of the mass is neither round nor oval.**

2.2. Data Pre-processing

The images of different sizes of feature training set are adjusted to 224*224 before the experiment to facilitate the training and reduce the amount of computation at the same time. And all the image values are normalized to improve the algorithm accuracy and accelerate the convergence speed of the algorithm. The images are randomly rotated by 90°, and then some of the images are vertically flipped and horizontally flipped. It helps to enhance the diversity of the data and reduce the overfitting caused by different orientations.

2.3. Network Model DSA-U++

We propose an improved deep learning model based on ResNet [1] and UNet++ [2] for feature classification of ultrasound breast images. Specifically, the DSA-U++ model first uses the R-AS module instead of the original encoder of U-Net++ for feature extraction, and the R-AS module employs ResNet50 as well as the ASPP module to adequately extract the multi-scale features of the image. Resnet is a network model that has been applied to visual tasks such as image classification and semantic segmentation. Its multiple stacking by a special residual module can avoid the situation of gradient vanishing or gradient explosion, but also can learn deep features in a short time. The R-AS module is shown in Fig 1.

The R-AS module contains 1 convolution block, 4 residual blocks, and 5 outputs. The convolution block is composed of 1 convolutional layer of size 7×7, batch normalization layer, activation layer, and 3×3 maximum pooling layer. The convolutional residual blocks are all composed of convolutional kernels of size 1 × 1, 3 × 3, and 1 × 1. The residual formula is shown in (1).

$$H(x) = F(x) + x \tag{1}$$

In the formula: $F(x)$ denotes the output of the convolutional layer and represents the output obtained after convolution; x is the input to the residual block, from the feature map of the previous layer; $H(x)$ are the outputs of the residual blocks.R-AS2, R-AS3, R-AS4, and R-AS5 are obtained from the outputs of 3, 4, 6, and 3 residual blocks, respectively.

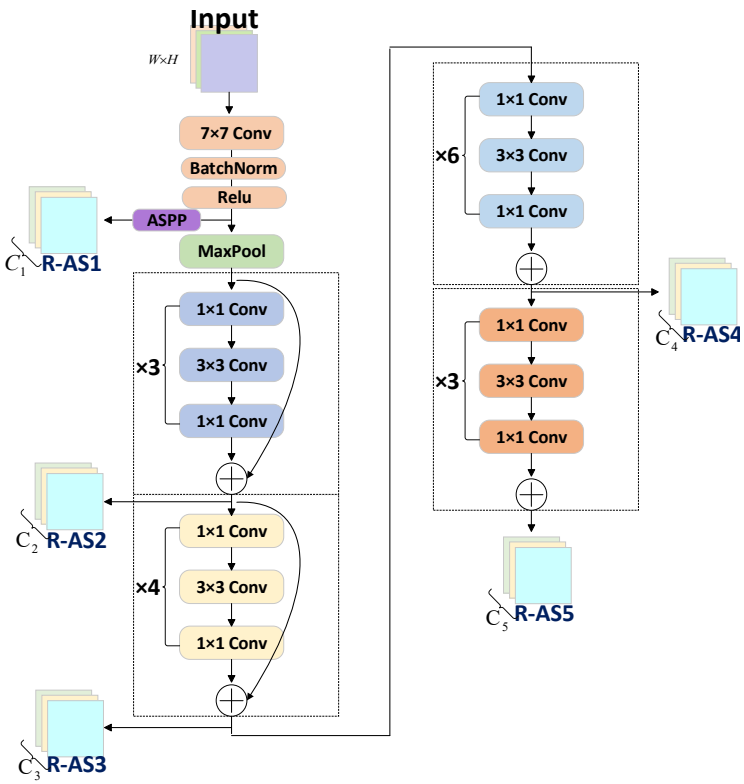


Fig 1. R-AS module

In our model DSA-U++, R-AS1 is rich in spatial detail information due to its inclusion. So, the output feature map R-AS1 is obtained after multi-scale feature extraction by applying the null-space convolutional pooling pyramid, together with the output of the four feature maps R-AS via four residual blocks, respectively, constitutes the encoder part of U-Net++. Subsequently, in the decoder part, we improve the network based on the U-Net++ architecture by replacing some of the traditional convolution operations with depth-separable convolutions, thus effectively reducing the number of model parameters and computational complexity while maintaining the feature expression capability. Five feature maps of different sizes R-AS replace the original encoder output of U-Net++, which is then decoded. $X^{i,j}$ is a convolutional unit. i denotes the i-th downsampling layer, j represents the jth convolutional layer in the current jump-connected layer, Suppose $x^{i,j}$ is the output of the convolution $X^{i,j}$. The output $x^{i,j}$ of each convolutional unit in Fig. 2 can be expressed by equation (2):

$$x^{i,j} = \begin{cases} H(D(x^{i-1,j})), & j = 0 \\ H([(x^{i,k})_{k=0}^{j-1}, U(x^{i+1,j-1})]), & j > 0 \end{cases} \quad (2)$$

In the formula: $D(\ )$ denotes the current node, $H(\ )$ denotes the convolution operation and the activation function; denotes the upsampling operation; $[\ ]$ denotes the feature map splicing operation. When $j = 0$, the node receives only inputs from the previous downsampled layer; when $j > 0$, the node receives $j+1$ inputs, including jump connections and the upsampled layer as inputs.

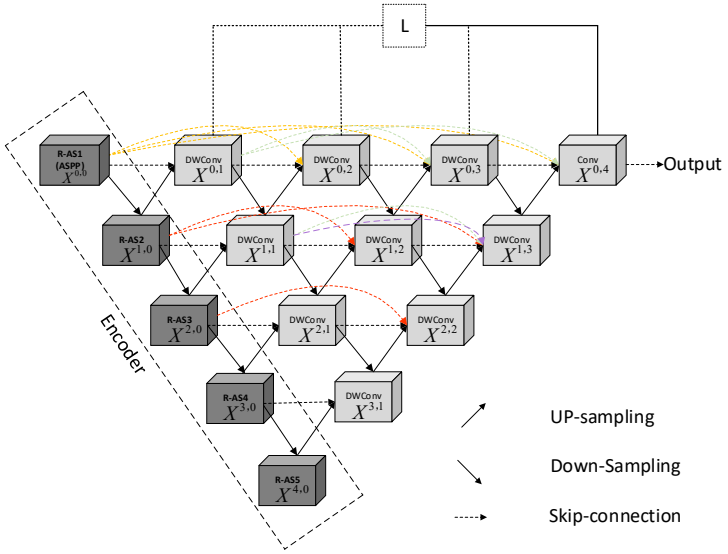


Fig 2. SA-U++ network model

We also replace some of the traditional convolution blocks with depth-separable convolutions in the DSA-U++ model.

Its effectively reduces the computational and parametric quantities of the model by splitting the standard convolution into depth convolution and point-by-point convolution, while retaining the important feature extraction capabilities. The output image of the decoder part goes through a fully connected layer for the classification task. With this improved design, especially in the case of ultrasound images where lesions are usually small and of low contrast, this optimization not only enhances the computational efficiency of the model, but also improves the classification accuracy to a small extent.

2.4. Depthwise Separable Convolution

Depthwise separable convolution was first proposed by howarda [1] and others. It is shown to be divided into two main processes, channel-by-channel convolution and point-by-point convolution. Channel-by-channel convolution differs from ordinary convolution in that its convolution kernel adopts a single-channel mode, where each channel of the input image is convolved to obtain an output feature map with the same number of channels as that of the input, which extracts the spatial features of the input image. Point-by-point convolution makes up for the shortcomings of channel-by-channel convolution that does not effectively utilize the feature information of different channels, using a 1*1 convolution kernel, and outputs a feature map after all channels are aggregated.

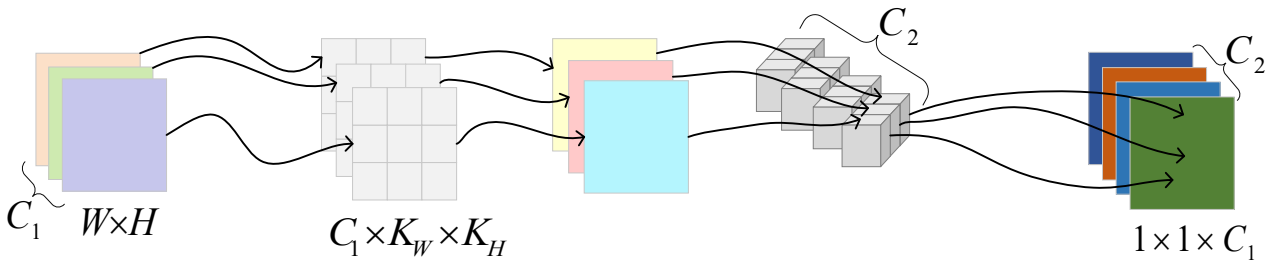


Fig 3. Depthwise Separable Convolution

First, channel-by-channel convolution convolves its input image using convolution kernels with the same number of channels C_1 respectively, and C_1 feature maps are obtained after convolution: $WHC_1K_WK_H$,However, the feature information between the channels was not utilized in the meantime. Thus, in point-by-point convolution, using $1\times1\times C_1$ convolution kernel for channel-by-channel convolution to get C_1 feature maps for inter-channel feature information fusion, to get C_2 number of channels of feature maps as the output, the computational volume of its convolution is: WHC_1C_2 . The computational effort of using depthwise separable convolution is: $WHC_1K_WK_H + WHC_1C_2$. The computational effort to obtain C_2 output feature maps using conventional convolution of input convolution with channel number C_1 and size K_WK_H is : $WHC_1K_WK_H$. The ratio of the computational effort of the two is : $\frac{1}{C_2}+\frac{1}{K_WK_H}$. The convolution kernel size K_WK_H used in this experiment is 3x3. It can be seen that the computational effort of depth-separable convolution is about 1/9th of that of conventional convolution.

2.5. Atrous Spatial Pyramid Pooling

The ASPP module is used to enhance the capability of convolutional neural networks in processing multi-scale information, especially in image segmentation tasks. ASPP was firstly applied in the DeepLab [1] family of models, aiming to enhance the model's ability to perceive features at different spatial scales through the introduction of null convolution at different scales and global average pooling. By capturing the contextual information of an image at different scales, the model's ability to distinguish between objects and backgrounds is improved, especially in images with complex structures and fuzzy boundaries.

Null convolution, i.e., inserting nulls in the convolution kernel to increase the size of the perceptual field without increasing the amount of computation or pooling operations. In the ASPP module of this experiment, cavity convolution with expansion rates of 6, 12, and 18 is used, which enables the model to perceive the features of the image at different scales and thus capture both local and global information. It is also capable of expanding the receptive field, thus increasing the model's ability to perceive larger regions.

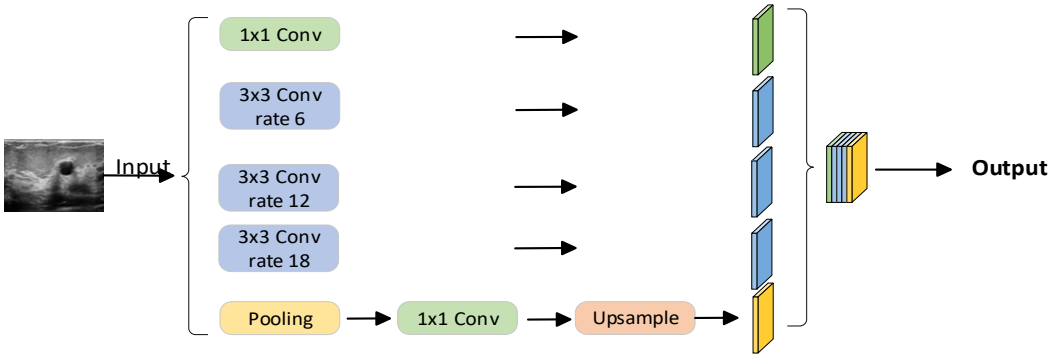


Fig 4. ASPP

The module consists of multiple parallel branches that capture features at different scales in the image by using 3x3 null convolution with different expansion rates. The 1x1 ordinary convolution, null convolution is spliced with the image obtained after global average pooling via 1x1 ordinary convolution and up adoption again. And further feature fusion

is performed by a 1x1 ordinary convolution to output multi-scale features. This enables the model to capture both local features and global information in the image, thus improving the robustness and accuracy of the model. As shown in Fig. 4.

3. Experiments and Results

3.1. Experimental Environment

This experiment experiments Pycharm as compiler, using Python-3.10.15 as compiler, the experimental framework is Pytorch. the hardware environment is 12th Gen Intel(R) Core (TM) i7-12700H 2.30 GHz, graphics card is NVIDIA GeForce RTX3060 Laptop GPU, 64-bit Windows operating system.

3.2. Evaluation Indicators

In this experiment, the accuracy Accuracy, F1 value was used to evaluate the result of ultrasound breast image feature recognition and the formula was calculated as follows:

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + \mathrm{F_N}} \tag{3}$$

$$Recall = \frac{T_P}{T_P + F_N} \tag{4}$$

$$\Pr ecesion = \frac{\mathrm{T_P}}{\mathrm{T_P} + \mathrm{F_P}} \tag{5}$$

$$F_1 = \frac{2 \times \Pr ecision \times \mathrm{Re}\, call}{\Pr ecision + \mathrm{Re}\, call} \tag{6}$$

In the formula: T_P refers to the number of correctly categorized positive samples, which are predicted to be positive and are actually positive; F_P refers to the number of negative samples incorrectly labeled as positive samples, which are actually negative samples predicted to be positive; T_N refers to the number of correctly categorized negative samples, which are predicted to be negative and are actually negative; and F_{N} refers to the number of positive samples incorrectly labeled as negative samples, which are actually positive samples predicted to be negative.

3.3. Experimental Results

In this experiment the DSA-U++ model is compared with known image classification models and modified image segmentation models. These models include Resnet34, U-Net, U-Net++, Vgg16 [1], and Msa2Net [2]. The experimental results are shown in the table.

Table 2. Accuracy, F1 value results for each model

model	Accuracy (%)				mAcc(%)	mF1(%)
	direction	**boundary**	**calcification**	shape		
U-Net	84.5	78.9	**87.7**	88.7	84.9	68.2
U-Net++	84.8	81.6	84.2	89.6	85.0	67.4
Vgg16	84.1	80.6	86.7	89.7	85.2	**68.8**
Msa2Net	84.1	80.4	83.3	89.1	84.2	**68.8**
DSA-U++	**85.4**↑	**79.8**↑	87.1	**89.8**↑	**85.5**↑	67.5

As shown in Table 2, the improved algorithm based on Resnet-U-Net++ for image classification is also effective and accurate. Compared to the common image classification algorithm Vgg16, there is also some improvement.

3.4. Conclusion

In this paper, we propose a deep learning model, DSA-U++, based on ResNet34 and U-Net++ improvement for feature classification of ultrasound breast images. The model employs ResNet34 as an encoder to efficiently extract multiscale features in the image, and also introduces the R-AS module designed by the convolutional pooling pyramid in the null space to enhance the representation of multiscale contextual information. Subsequently, in the decoder part, based on the improved DAS-U++ model, we replace part of the traditional convolutional blocks with depth-separable convolutions, which not only reduces the number of model parameters and the computational complexity while maintaining the feature expressiveness, but also preserves the detailed information of the image more adequately. The dense hopping connections of U-Net++ also help to fuse the different levels of features, enabling the model to can utilize more contextual information.

With the above improved design, the DSA-U++ model is able to capture subtle lesion features in ultrasound breast images more comprehensively and achieve accurate classification based on multi-level feature fusion. The experimental results show that the model also improves in classification accuracy compared to the traditional model, which verifies the effectiveness of the proposed method in medical image feature extraction and classification tasks.

Acknowledgments

Funded Project: 2025 Student Innovation and Entrepreneurship Training Program Project (Feature Recognition of Ultrasound Breast Images Based on UNet++).

References

[1] J. Ferlay, M. Colombet, I. Soerjomataram, D.M. Parkin, M. Piñeros, A. Znaor, F. Bray, Cancer statistics for the year 2020: An overview, International journal of cancer (2021).

[2] I. Soerjomataram, F.J.N.r.C.o. Bray, Planning for tomorrow: global cancer incidence and the role of prevention 2020–2070, 18(10) (2021) 663-672.

[3] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries, CA: a cancer journal for clinicians 71(3) (2021) 209-249.

[4] R. Guo, G. Lu, B. Qin, B. Fei, Ultrasound Imaging Technologies for Breast Cancer Detection and Management: A Review, Ultrasound in Medicine & Biology 44(1) (2018) 37-70.

[5] S.C. Chen, Y.C. Cheung, C.H. Su, M.F. Chen, T.L. Hwang, S.J.U.i.O. Hsueh, G.T.O.J.o.t.I.S.o.U.i. Obstetrics, Gynecology, Analysis of sonographic features for the differentiation of

benign and malignant breast tumors of different sizes, 23(2) (2004) 188-193.

[6] S.T. Chen, Y.H. Hsiao, Y.L. Huang, S.J. Kuo, H.S. Tseng, H.K. Wu, D.R. Chen, Comparative analysis of logistic regression, support vector machine and artificial neural network for the differential diagnosis of benign and malignant solid breast tumors by the use of three-dimensional power Doppler imaging, Korean journal of radiology 10(5) (2009) 464-71.

[7] J. Redmon, S.K. Divvala, R.B. Girshick, A.J.I.C.o.C.V. Farhadi, P. Recognition, You Only Look Once: Unified, Real-Time Object Detection, (2015) 779-788.

[8] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A.C. Courville, Y. Bengio, Generative Adversarial Nets, Neural Information Processing Systems, 2014.

[9] A. Vaswani, N.M. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention is All you Need, Neural Information Processing Systems, 2017.

[10] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, North American Chapter of the Association for Computational Linguistics, 2019.

[11] T. Steifer, M. Lewandowski, Ultrasound tissue characterization based on the Lempel–Ziv complexity with application to breast lesion classification, Biomedical Signal Processing and Control 51 (2019) 235-242.

[12] O. Ronneberger, P. Fischer, T.J.A. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, abs/1505. 04597 (2015).

[13] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation, International Conference on Medical Image Computing and Computer-Assisted Intervention, 2016.

[14] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778.

[15] Z. Zhou, M.M.R. Siddiquee, N. Tajbakhsh, J.J.D.L.i.M.I.A. Liang, D. Multimodal Learning for Clinical Decision Support : 4th International Workshop, M.-C. 8th International Workshop, held in conjunction with MICCAI , Granada, Spain, S... U-Net++: A Nested U-Net Architecture for Medical Image Segmentation, 11045 (2018) 3-11.

[16] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H.J.A. Adam, MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, abs/1704.04861 (2017).

[17] L.-C. Chen, G. Papandreou, I. Kokkinos, K.P. Murphy, A.L.J.I.T.o.P.A. Yuille, M. Intelligence, DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs, 40 (2016) 834-848.

[18] K. Simonyan, A.J.C. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, abs/1409.1556 (2014).

[19] S.G. Kolahi, S.K. Chaharsooghi, T. Khatibi, A. Bozorgpour, R. Azad, M. Heidari, I. Hacihaliloglu, D. Merhof, MSA^{2}Net: Multi-scale Adaptive Attention-guided Network for Medical Image Segmentation, 2024.