

Thyroid Cancer Ultrasound Image Segmentation Based on UNet Dimensional Variants and Transfer Learning GANs

Minghao Liu

*Faculty of Computer Science and Technology, Qilu University of Technology, Jinan, China
18005436269@163.com*

Abstract. UNet is a mainstream technique in medical image segmentation, and it is one of the most important semantic segmentation frameworks based on CNNs. With ongoing development, various UNet-related architectures and hybrid models have emerged, such as UNet++, Attention UNet, 3D UNet, and Res UNet. This paper focuses on the basic conceptual principles of UNet, the advantages and disadvantages of its variants on the performance of cancer image segmentation. Furthermore, the study explores the impact of transfer learning-based GAN on the performance of thyroid cancer image segmentation. Specifically, we examine the effects of different UNet dimensional variants (e.g., 2D, 2.5D, 3D) on segmentation performance, as well as the enhancement introduced by GAN-generated synthetic data through transfer learning. Experimental results show that the 2.5D UNet achieved the best performance on a dataset of 800 images, with a Test Dice score of 0.628 and a Jaccard index of 0.593. When the training set consisted of 600 original images and 400 GAN-generated synthetic images, the model trained with these augmented data achieved the best overall performance, with a Test Dice of 0.628 and a Jaccard of 0.525.

Keywords: image segmentation, UNet, GANs, Cancer

1. Introduction

With the improvement of medical treatment and living standards, the average life expectancy of people has been increasing, and with it, the problem of cancer has been aggravated. Limited by the level of treatment in the field of cancer, people tend to advocate prevention rather than treatment, i.e., early detection and advanced treatment are the most ideal situations. However, as cancer is still in its early stage, the coverage to the extent of the lesion is not obvious; doctors can only rely on their experience to make judgements in the absence of sufficient evidence, which may lead to misdiagnosis. At this point, we can adopt two approaches to improve the present situation. First, we can improve the technology of CT, MRI and ultrasonography, so that the films taken can provide more sufficient evidence on whether or not there is cancer. Secondly, we can improve the level of doctors, so that they can judge more accurately whether they have cancer or not under the limited conditions provided by the current technology. As a result, the field of medical image analysis has grown rapidly in the last 20 years [1].

Medical image analysis is increasingly being applied in clinical diagnosis. Although it can't be completely independent of machine diagnosis, which may involve issues such as technical

immaturity and accountability for accidents, it can already be used as an excellent assistant for doctors to participate in medical diagnosis. Medical image segmentation, as one of the most important aspects, can collect important features from the image and calculate the Dice coefficient, Jaccard index, and other performance indicators, which intuitively provide doctors with comprehensive evidence for analysis [2, 3].

With the emergence of machine learning, it has been increasingly developed and refined in the field of image segmentation. Initially, techniques such as edge detection, texture feature extraction, and morphological filtering were primarily used to automatically extract features. However, these methods were found to be too simplistic and limited in handling more complex tasks. After that, Johns Hopkins University Hubel and others proposed the idea of hierarchical processing. By adding hidden layers to a monotonic network, they enabled it to address more complex problems and automatically learn to extract multiple features across several hidden layers. This marked the advent of deep learning [4, 5].

The UNet network is one of the most representative segmentation models in deep learning, proposed by Ronneberger et al. UNet is a deep learning architecture for biomedical image segmentation, featuring a U-shaped structure that consists of contraction and expansion paths. The contraction path reduces the spatial dimension and increases the feature depth by extracting features through convolution and pooling, while the expansion path recovers the spatial dimension through up-sampling and convolution, and improves the segmentation accuracy by combining the feature maps of the contraction path with the use of hopping connections. The symmetric structure and the end-to-end training of UNet have made it a classic model with excellent performance in medical image segmentation [6].

This article focuses on the problems and optimization methods of UNet in cancer ultrasound imaging segmentation applications. We primarily tested two optimization approaches. The first one is to use a higher-dimensional UNet model to enable the model to learn more comprehensive and detailed features. The experiment compares the performance differences between 2D UNet and 2.5D UNet. The second is to use the combination of transfer learning and GAN to transfer the weights of the GAN model for ultrasound images of brain cancer to improve GAN stability and accuracy. This method also generates excellent ultrasound images of thyroid cancer even with a small training set.

2. Impact of UNet model dimensional variants on cancer image segmentation performance

2.1. Segmentation performance of 2D Unet 2.5D Unet

In most cases, increasing the dimensions of the UNet model leads to performance improvements, but this comes with higher memory requirements and increased computational complexity. For different application scenarios, we need to take the most appropriate way to deal with the situation, rather than just considering the performance.

Su Yang et al. have found that 3D CNN is the best, 2.5D CNN is the second best, and 2D CNN is relatively the worst in the image segmentation work, with a more significant performance difference among the three [7]. This proves the feasibility of increasing the dimensionality for improving the accuracy of image segmentation. To be more specific about the application of the UNet model in cancer PET/CT image scenario, Mahdi et al. investigated the application of 2D, 2.5D, and 3D UNet in head and neck tumors in dual Positron Emission Tomography/Computed Tomography (PET/CT) imaging. The results similarly demonstrated that 3D UNet achieved the best segmentation performance, followed by 2.5D UNet, while 2D UNet performed the worst [8].

These cases demonstrate that, whether it is the generalized CNN for normal images or specifically the segmentation effect of UNet for cancer PET/CT images, the performance improves more significantly as the model dimensions go from 2D, 2.5D to 3D.

2.2. Experiment 1: Effect of UNet dimensions on segmentation performance of thyroid cancer ultrasound images

With the continuous development of medical standards, ultrasound imaging has gradually become one of the mainstream modalities for tumor detection, which is efficiently applied in the diagnosis of thyroid cancer. Yang Guang et al. [9] made a study on the segmentation of thyroid tumor ultrasound images, demonstrating the feasibility and potential of this approach. However, existing studies have not yet explored whether the segmentation performance of thyroid tumor ultrasound imaging can be further optimized by increasing the dimensionality of the UNet model. The present experiment aims to address this gap.

In this study, the ultrasound imaging videos of a thyroid cancer patient were used, and individual frames were extracted and treated as separate ultrasound images. The training set of 2D UNet was used with 600 and 800 images, respectively. 2.5D, on the other hand, was in the form of sampled context and the training set was used in the three-channel form with 600 and 800 images, respectively. The performance metrics for the two sets of experiments were as follows:

Table 1. 2D UNet performance indicators

Training Set Size	Number of Epochs	Training Time	Train_loss	Val_loss	Train_dice	Val_dice	Test_dice	Jaccard
600	103	10min	0.197	0.392	0.662	0.380	0.404	0.405
800	131	12min	0.117	0.373	0.731	0.401	0.569	0.417

Table 2. 2.5D UNet performance indicators

Training Set Size	Number of Epochs	Training Time	Train_loss	Val_loss	Train_dice	Val_dice	Test_dice	Jaccard
600	84	57min	0.071	0.230	0.8771	0.5851	0.610	0.528
800	99	69min	0.050	0.199	0.9056	0.6391	0.708	0.593

Table1 and Table 2 record the performance metrics of 2D UNet and 2.5D UNet. When the training set size was 600, the 2D UNet achieved a Test Dice coefficient of 0.404 and a Jaccard index of 0.405, while the 2.5D UNet achieved significantly higher scores of 0.610 and 0.528, respectively. With a training set of 800 images, the 2D UNet reached a Test Dice of 0.569 and a Jaccard of 0.417, whereas the 2.5D UNet further improved to 0.708 and 0.593, respectively. It can be seen from the performance metrics of the two tests that the segmentation performance of the thyroid cancer ultrasound image of 2.5D UNet is significantly better than that of 2D UNet, and the segmented thyroid cancers have higher overlap with the real locations.

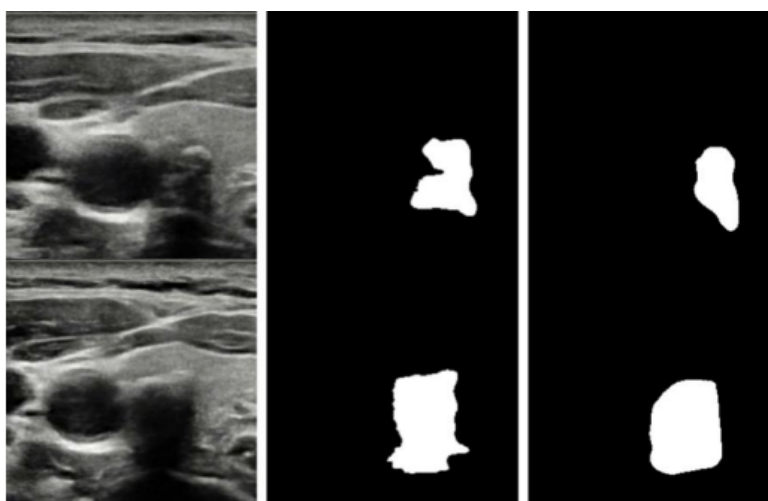


Figure 1. 2D UNet split effect

(Ultrasound image on the left, mask image labelled by the doctor in the middle, segmentation of UNet on the right)

Figure 1 and Figure 2 are the segmentation diagrams of 2D UNet and 2.5D UNet, respectively. It can be seen that the approximate positions of the segmentation of 2D UNet and 2.5D UNet are more accurate, but the edge contour of 2.5D UNet achieves more accurate segmentation and the overall shape is more in line with the real data compared with 2D UNet.

According to the performance index and the segmentation effect graph, it can be seen that the segmentation effect of 2.5D UNet is better than that of 2D UNet.

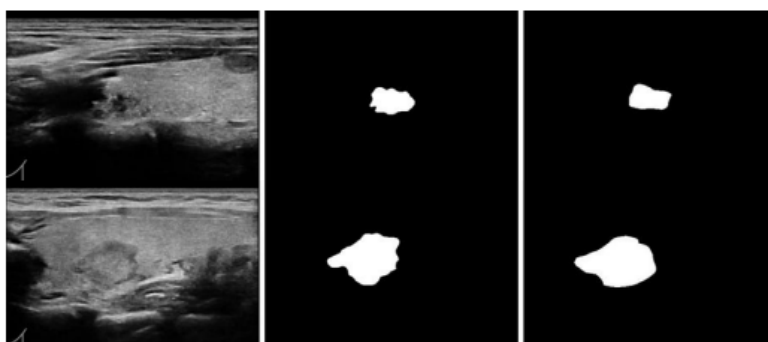


Figure 2. 2.5D UNet split effect

(Ultrasound image on the left, mask image labelled by the doctor in the middle, segmentation of UNet on the right)

3. Impact of transfer learning-based GAN on the segmentation performance of thyroid cancer ultrasound imaging

3.1. Application of GAN in medical image generation

In the medical domain, ethical, privacy, and regulatory constraints often limit access to large, high-quality datasets. This challenge is particularly pronounced in tumor image segmentation, where acquiring sufficient annotated medical images is difficult. UNet is a specific kind of CNN that also falls under the category of deep learning, which leads to the fact that it requires a large amount of data for training purposes. In order to better solve a particular problem, we always need to look for

different imaging of various tumors, which makes it difficult to get effective and sufficient data to use as a training set.

In order to better address the above problems, Goodfellow first proposed the concept of GAN model. GAN is an innovative deep learning model which consists of two parts: generator and discriminator. The generator is responsible for generating new data, while the discriminator is responsible for determining whether the data is real or generated. Through adversarial training between these two networks, GAN is able to continuously improve the quality of the generated data, making it closer and closer to the real data. This model has a wide range of applications in the fields of image generation, style migration, data enhancement, etc [10-12].

One of the most extensively studied imaging modalities is computed tomography (CT). In this context, Zhen Hu introduced the idea of GAN-DAUNet in his research [13], and the combination of the attention mechanism and GAN achieved good segmentation results in processing liver CT images. The synthesis of CT images by MR studied by Hai-Tao Wu [14] also achieved excellent results. When segmenting lung nodules [15], Guitang Wang also used GAN to generate expanded samples, which verified that the use of generative adversarial networks to generate samples can improve the segmentation performance very well. Then there is ultrasound. In the segmentation of brachial plexus nerve ultrasound imaging studied by Yuchao Lu [16], the combination of GAN and UNet also achieved relatively satisfactory results. On synthetic breast grey scale ultrasound images [17], Li Yang et al. proposed a method for generating elastography images that closely resemble real elastography scans, which can effectively assist clinicians in diagnosis and treatment. These findings suggest that GANs can effectively compensate for the limitations of small training datasets and contribute to improving the segmentation performance of UNet-based models.

However, the most significant problem with the GAN model is that the data requirement for training is large and extremely unstable. We adopt the idea of transfer learning to accelerate the convergence of stable models by using the weights of pre-trained models to reduce the training set demand and improve the training effect. Transfer learning is a branch of machine learning that refers to the application of existing models to new but related problems. Its core idea is to use existing knowledge to solve new problems, which can save a lot of data collection and model training time and improve learning efficiency.

3.2. Experiment 2: GAN models combined with transfer learning: improving the segmentation performance of ultrasound images of thyroid cancer

We use the brain tumor ultrasound imaging generation model combined with transfer learning, using its pre-trained weights as the initial parameters for our own thyroid tumor model. For the same tumor-related model, the tumors have good similarities with each other, such as edge ambiguity, internal heterogeneity, shape irregularity, etc., and the brain tumor generation model has meaningful reference value for the thyroid tumor generation model.

The medigan open model was used for this experiment [18]. We selected the model in which ultrasound imaging was generated and set up two control groups. The first group was randomly initialized with weights and the second group used brain tumor model weights as initialization weights. Both groups used data from 200 ultrasound images of thyroid tumors in patients as the training set. After training, then 400 simulated lesion images were generated separately as supplementary data for the 2D UNet model training set.

It can be seen that Figure 3 and Figure 4 are the generation effect graphs of GAN before and after transfer, respectively.

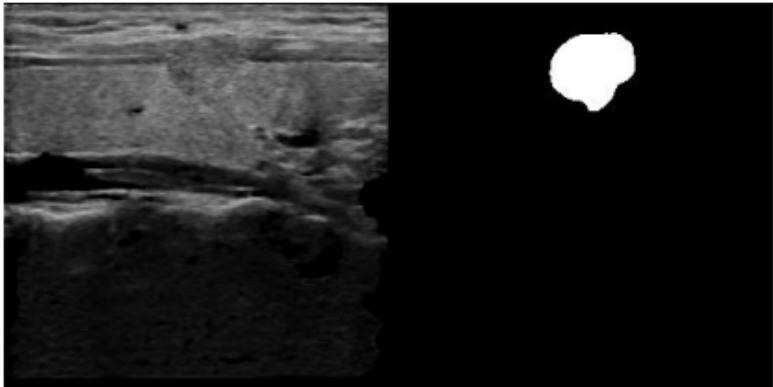


Figure 3. Generation of renderings before transfer learning

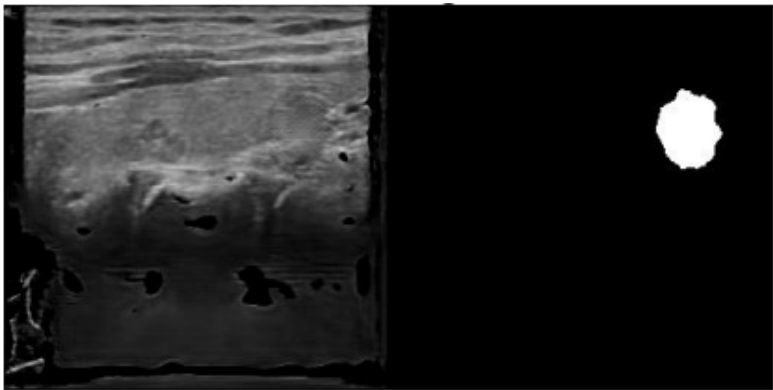


Figure 4. Generate effect diagram after transfer learning

Both sets of GAN generation effect images are reasonable to generate the tumor correctly into the tissue part. The edge contour of the tumor generated by the unmigrated GAN is more rounded, while the edge contour of the tumor generated by the post-migrated GAN is more angular, and the learning of the tumor contour is more refined. The performance results of the two experimental groups are presented in Table 3 and Table 4.

Table 3. UNet performance without transfer learning

Training Set Size	Number of Epochs	Training Time	Train_loss	Val_loss	Train_dice	Val_dice	Test_dice	Jaccard
600: 200	148	24min	0.184	0.322	0.705	0.46	0.479	0.407
600: 400	73	15min	0.255	0.342	0.685	0.506	0.509	0.433

(The size of the training set in Table, preceded by the number of original datasets and followed by the number of GAN-generated datasets)

Table 4. UNet performance after migration

Training Set Size	Number of Epochs	Training Time	Train_loss	Val_loss	Train_dice	Val_dice	Test_dice	Jaccard
600: 200	69	11min	0.1962	0.2968	0.691	0.527	0.586	0.481
600: 400	61	11min	0.1087	0.2007	0.783	0.597	0.628	0.525

(The size of the training set in Table, preceded by the number of original datasets and followed by the number of GAN-generated datasets)

When the training set size is 600:200, the Test-dice and Jaccard of UNet are 0.479 and 0.407, respectively, when the unmigrated GAN-generated images are used as the training set, and the Test-dice and Jaccard of UNet are 0.586 and 0.481, respectively when the migrated GAN-generated images are used as the training set. When the training set size is 600:400, the Test-dice and Jaccard of UNet are 0.509 and 0.433 for the unmigrated GAN-generated images as the training set, and 0.628 and 0.525 for the migrated GAN-generated images as the training set. Through the UNet's segmentation performance, we can see that the performance of the migrated GAN model for the segmentation task is significantly better than the pre-migration GAN model. This proves the feasibility of combining transfer learning with GAN, which well improves the accuracy and effectiveness of the data generated by GAN, and the generated graphs are put into the training set to better improve the segmentation effect of UNet.

4. Conclusion

With a test-dice of 0.628 and a Jaccard of 0.593, the 2.5D UNet significantly outperforms the 2D UNet on a training set of 800 images. From this, we can see that with the increase of dimensionality, the performance of the UNet model for segmentation of ultrasound imaging of thyroid cancer can be effectively improved. When the migrated GAN model generates a training set of 400 images supplemented with 600 original images, the Test-dice is 0.708 and the Jaccard is 0.525, and the performance is better than that of the pre-migration GAN model. Finally, we can infer that transfer learning can optimize the performance of the GAN model to generate diverse ultrasound imaging of thyroid cancer more consistently and accurately. In summary, this study demonstrates that increasing UNet dimensionality improves segmentation accuracy, and combining GAN with transfer learning further enhances the effectiveness of data augmentation and model training in the context of segmenting ultrasound images of thyroid cancer.

References

- [1] Huaxin YU, & Liwen HE. (2024). Research on Image Segmentation Based on UNet. *Software Engineering*, 27, 50–53. Doi : 10.19644/j.cnki.issn2096-1472.2024.010.010 [2] XU Guangxian, FENG Chun, & MA Fei. (2023). Review of Medical Image Segmentation Based on UNet. *Journal of Frontiers of Computer Science and Technology*, 17, 1776–1792. doi: 10.3778/j.issn.1673-9418.2301044
- [3] Zhao, Bing. (2024). Research on segmentation algorithm of eye ultrasound image based on improved UNet model. *Modeling and Simulation*, 13, 5808.
- [4] Juanxiu Tian, Guocai Liu, Shanshan Gu, Zhongjian Ju, Jinguang Liu, & Dongdong Gu. (2018). Research and challenges of deep learning methods for medical image analysis. *Journal of Automation*, 44(3), 401-424.
- [5] Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160(1), 106.
- [6] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [C]// LNCS 9351: Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Oct 5-9, 2015. Cham: Springer, 2015: 234241.
- [7] Su Yang, Jong Soo Jeong, Dahyun Song et al. Comparison of 2D, 2.5D, and 3D segmentation networks for mandibular canals in CBCT images: A study on public and external datasets, 21 October 2024, PREPRINT (Version 1) available at Research Square [<https://doi.org/10.21203/rs.3.rs-5035833/v1>]
- [8] Mahdi, M. A., Ahamad, S., Saad, S. A., Dafhalla, A., Alqushaibi, A., & Qureshi, R. (2024). Segmentation of Head and Neck Tumors Using Dual PET/CT Imaging: Comparative Analysis of 2D, 2.5 D, and 3D Approaches Using UNet Transformer. *CMES-Computer Modeling in Engineering & Sciences*, 141(3).
- [9] Yang Guang, Wen He, Jiajun Wu, Mingchang Zhao, Yukang Zhang, Fang Wan. Clinical application of deep learning based ultrasound segmentation of thyroid nodules [J/OL]. *Chinese Journal of Medical Ultrasound* (Electronic Edition), 2022, 19(03): 206-211.

- [10] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, 27.
- [11] Osuala, R., Skorupko, G., Lazrak, N., Garrucho, L., García, E., Joshi, S., ... & Lekadir, K. (2023). medigan: a Python library of pretrained generative models for medical image synthesis. *Journal of Medical Imaging*, 10(6), 061403-061403.
- [12] E. Kim, H. -H. Cho, J. Kwon, Y. -T. Oh, E. S. Ko and H. Park, "Tumor-Attentive Segmentation-Guided GAN for Synthesizing Breast Contrast-Enhanced MRI Without Contrast Agents, " in *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 11, pp. 32-43, 2023, doi: 10.1109/JTEHM.2022.3221918.
- [13] Zhen Hu. (2024). Research on GAN-based segmentation method for liver CT images. Southwest University of Science and Technology. doi: 10.27415/d.cnki.gxngc.2024.000369.
- [14] Haitao Wu. (2020). MR to CT Synthesis Based on Cycle-Consistent Generative Adversarial Network. University of Chinese Academy of Sciences(Shenzhen Institutes of Advanced Technolog), 7. doi: 10.27822/d.cnki.gszxj.2020.000111.
- [15] Guitang Wang, Zhenzhe Lin, Qinshen Fu, Jingran Wang, & Guojie Lu. (2020). Joint generative adversarial network model for classification of benign and malignant pulmonary nodules. *Chinese Journal of Scientific Instrument*, 11, 188–197. doi: 10.19650/j.cnki.cjsi.J2006741.
- [16] Yuchao Lv. (2024). Deep learning-based segmentation and reconstruction of plane-wave B-Mode ultrasound images. Qingdao University of Science & Technology. doi: 10.27264/d.cnki.gqdhc.2023.000011
- [17] Yang Li, Jinyu Cai, Xiaozhi Dang, Wanying Chang, Yan Ju, Yi Gao, & Hongping Song. (2024). Application value of a deep learning-based model for generating strain elastography images using breast grayscale ultrasound images. *Chinese Journal of Ultrasound in Medical Sciences (Electronic Edition)*, 21(06), 563-570.
- [18] Mansheng Long, Chunjuan Ouyang, Huan Liu, & Qing Fu. (2018). Image recognition of *Camellia oleifera* diseases based on convolutional neural network & transfer learning. *Transactions of the Chinese Society of Agricultural Engineering*, 34(18).