Proceedings of CONF-CIAP 2026 Symposium: Applied Mathematics and Statistics
DOI: 10.54254/2753-8818/2026.CH31301

Methods and Challenges of U-Net in CT Segmentation of Pulmonary Micronodules

Xilin Guo

College of Smart Agriculture (College of Artificial Intelligence), Nanjing Agricultural University, Nanjing, China
g13167547604@outlook.com

Abstract. Pulmonary micronodules often present with weak contrast, blurred boundaries, and easy adhesion to blood vessels or pleura in Computed Tomography (CT), which leads to missed detection and missegmentation in automatic segmentation, affecting the consistency of screening and follow-up evaluation. This paper reviews the research progress of U-Net and its improved structures in the segmentation of pulmonary micronodules. It focuses on summarizing technical routes such as multi-scale feature fusion, attention mechanism, three-dimensional convolution, and convolutional Transformer hybrid structure, and summarizes and compares the performance of common indicators of related methods on datasets such as DCC-IDRI and LUNA16. Overall, multi-scale dense connection, spatial-channel attention, and long-range dependency modeling are helpful to improve the stability of edge characterization and segmentation of small nodules, but they are still constrained by factors such as weak features, background complexity, and cross-device domain offset. In the future, the robustness and generalization ability in multi-center scenarios can be further enhanced from directions such as lightweight 3D modeling, multi-modal fusion, semi-supervised learning, and explainability mechanisms.

Keywords: Pulmonary Micronodules, U-Net, Attention Mechanism, Three-Dimensional Segmentation, Transformer

1. Introduction

Lung cancer has long ranked among the top in terms of incidence and mortality among malignant tumors in China and globally. A considerable number of cases present in the form of pulmonary micronodules in the early stage. With the popularization of low-dose spiral Computed Tomography (CT) (LDCT) in lung cancer screening, a large number of tiny nodules with diameters of less than a few millimeters and adhering to blood vessels or pleural structures have been detected. Such lesions present weak contrast and blurred boundaries. Manual identification and delineation are not only time-consuming but also prone to inter-observer differences, posing challenges to precise quantitative screening, follow-up, and risk assessment. Therefore, developing a stable and highly accurate automatic segmentation method for micronodules is of great significance for improving the detection rate of early-stage lung cancer and promoting the clinical application of intelligent diagnostic systems.

In recent years, deep learning has driven the rapid development of pulmonary nodule segmentation technology, among which the encoder-decoding structure represented by U-Net has become the main framework. U-Net relies on skip connections and simultaneously utilizes local details and high-level semantic information to achieve high segmentation accuracy under limited annotation conditions [1,2]. Its subsequent improvements, such as UNet++, enhance the feature expression ability through dense multi-scale connections and perform more stably in complex backgrounds and small-volume nodules [3]. The dual-attention U-Net utilizes spatial and channel attention mechanisms to enhance the network's recognition ability for vascular attachments, parappleural areas, and regions with complex textures, achieving a higher Dice coefficient and recall rate on LUNA16 [4]. In addition, researchers have combined convolution with Transformer to model remote dependencies, introduced a detection-segmentation two-stage structure to improve the detection rate of small targets, and alleviated the problems of insufficient labeling and domain differences through semi-supervised or knowledge distillation strategies [5-7]. These developments have jointly enhanced the robustness and practicality of automatic segmentation of pulmonary nodules.

On this basis, this paper systematically reviews and analyzes the application progress of U-Net and its improved structures in the segmentation of pulmonary micronodules, and focuses on summarizing the advantages and limitations of technical routes such as multi-scale feature aggregation, attention mechanism, three-dimensional convolution, and convolution-Transformer hybrid structure. The research aims to enhance the segmentation accuracy, contour integrity, and cross-case robustness of pulmonary micronodules while maintaining a controllable computational load of the model, providing a more reliable quantitative tool for lung cancer screening and follow-up evaluation.

2. Theoretical basis

Pulmonary micronodules usually present characteristics such as small volume, insignificant difference in density from the surrounding lung parenchyma, and blurred boundaries on CT images. These factors make the automatic segmentation task itself highly challenging. Shen Quanyou et al. pointed out that the size, location, and shape of pulmonary nodules vary significantly, and they are often accompanied by blurred boundaries and proximity to structures such as blood vessels. These factors increase the difficulty of the segmentation task and make the model prone to missegmentation or missed segmentation [2]. These factors are also the direct driving forces behind the wide application of deep learning methods in the segmentation of micronodules.

2.1. Basic structure of U-Net

The U-Net proposed by Ronneberger et al. has been widely applied in medical image segmentation [1]. U-Net is composed of symmetrical encoders and decoders: the encoding end gradually expands the receptive field and extracts high-level semantic information through multi-layer convolution and downsampling; The decoding end restores the spatial resolution layer by layer through upsampling. The core lies in multi-level skip connections, which concatenate the feature mapping of the corresponding layer at the encoding end with the feature concatenation at the decoding end, enabling the network to retain fine-grained edge and texture information while also possessing strong semantic expression capabilities. Under the condition of limited labeled samples, U-Net can still obtain stable segmentation results relying on this structure and has obvious advantages in overlapping indicators compared with traditional segmentation methods [1,2].

2.2. U-Net improved structure

Under the U-Net framework, researchers have proposed a variety of improvement ideas for problems such as small nodule volume, strong background interference, and insufficient generalization caused by data differences across devices.

2.2.1. UNet++ and multi-scale feature fusion

The UNet++ proposed by Zhou et al. alleviates the semantic differences between the encoder and the decoder by reorganizing the skip connections and enhances cross-scale information fusion through dense connections [3]. Specifically, UNet++ introduces a series of intermediate nodes on the basis of the original U-shaped structure. Through step-by-step convolution and feature fusion, the shallow and deep features are fully aligned before entering the decoding end, thereby reducing the optimization difficulties caused by the "semantic gap". This structure can more accurately characterize the edge and internal structure of nodules when dealing with small-sized or complex-background nodules, which is conducive to improving indicators such as Dice and mIoU [2,3].

2.2.2. Attention-enhanced U-Net

Oktay et al. proposed the Attention U-Net, introducing attention gating in skip connections to suppress background responses, laying the foundation for subsequent attentional enhanced segmentation [8]. In the direction of the attention mechanism, Hou Yingzhu et al. introduced spatial and channel dual attention modules in U-Net, enabling the network to provide higher responses to key structures in parapapular, vascular attachment, and textured complex areas, thereby reducing the phenomenon of missegmentation [4]. Channel attention adaptively adjusts the weights by learning the importance of different feature channels, while spatial attention highlights the potential lesion areas within the planar range. The combined effect of the two enables the network to be more "focused" on the tiny nodules themselves in complex backgrounds, enhancing the segmentation stability on datasets such as LUNA16 [4].

2.3. Dimensional design: two-dimensional and three-dimensional U-Net

In terms of dimension design, the two-dimensional U-Net focuses more on the local context of a single layer or adjacent slices. Generally, 2D convolution is used to process axial or coronal images, which have the advantages of simple implementation and relatively low resource requirements. In contrast, the three-dimensional U-Net directly learns the overall spatial morphology of nodules through 3D convolution, which can better utilize the continuity of volume data. In lesions with blurred boundaries, such as ground-glass nodules or pleural nodules, segmentation results that are closer to manual labeling are often obtained [2]. However, three-dimensional structures have significantly higher computational load and memory overhead than two-dimensional models, which limits their deployment in conventional clinical environments. This is one of the reasons why subsequent lightweight 3D networks and ROI trimming strategies have been proposed [2].

2.4. Convolutional - transformer hybrid structure

Recent studies have gradually explored the structural design that combines convolution and Transformer to simultaneously take into account local texture modeling and long-range dependent expression. Among them, the S3TU-Net proposed by Wu et al. combines structural convolution, the

Transformer module, and superpixel features. It enhances the interaction ability of cross-scale features through large receptive field convolution and self-attention mechanism, and uses superpixels to partition and constrain the feature aggregation region, thereby improving the modeling effect of long-distance dependencies [6]. In the segmentation of nodules with low contrast, small volume, or complex background, this type of structure achieves a more stable performance compared to the traditional U-Net, especially having advantages in contour integrity and volume consistency [6].

The existing improvement methods of U-Net mainly focus on multi-scale feature fusion, attention enhancement, and 3D/hybrid structure modeling, providing effective ideas for weak contrast and small target segmentation. However, there is still a problem of insufficient stability in scenarios such as vascular adhesion, complex backgrounds, and cross-center domain shifts. The next section will further explain this in combination with specific applications and cases.

3. Applications and cases

U-Net and its improved structures have been widely applied in the field of automatic segmentation of pulmonary nodules. Relevant studies cover multiple directions, such as two-dimensional and three-dimensional segmentation processes, attention-enhanced models, and convolution-Transformer hybrid structures.

3.1. Two-dimensional U-Net segmentation process and performance characteristics

From the perspective of the actual workflow, a relatively common approach remains a two-dimensional U-Net: usually, single-layer or adjacent CT slices are taken as input, each layer is independently segmented, and then volume reconstruction is carried out in three-dimensional space [2]. Studies based on datasets such as LIDC-IDRI have shown that U-Nets in two-dimensional or multi-task forms can maintain relatively stable Dice coefficients and volume errors in most cases. Especially for solid nodules with larger diameters, segmentation results close to those of manual delineation can be obtained [2]. However, due to their weak contrast, small diameter, and frequent adhesion to vascular structures, micronodules are more prone to incomplete contours or discontinuous three-dimensional reconstructions, resulting in their overall segmentation effect generally being lower than that of larger lesions.

3.2. Application of attention-enhanced U-Net in complex backgrounds

To improve the segmentation stability in complex backgrounds, some studies have added spatial attention and channel attention modules in the U-Net to enhance the model's sensitivity to suspicious regions [4]. The dual-attention U-Net was verified on the LUNA16 dataset for locations such as the peripleural area and vascular cross. The results showed that in regions with obvious interference factors, the nodule contour extraction of the model was more stable, and there was a significant improvement in the Dice coefficient and recall rate compared with the original U-Net [4]. This indicates that in the case of weak image signals and complex background structures, the attention mechanism can effectively improve the recognition rate of tiny nodules and edge-blurred lesions.

3.3. 3D U-Net and lightweight strategy

For cases of ground-glass nodules or more irregular shapes, the three-dimensional U-Net, by taking advantage of the spatial continuity of volume data, can usually obtain segmentation results that are closer to manual labeling, especially having advantages in three-dimensional volume consistency and shape preservation [2]. However, the computational and video memory overhead brought by three-dimensional convolution is significantly higher than that of two-dimensional methods, and there are certain difficulties in deployment in large-scale screening and conventional workstation environments. Therefore, some works combine lightweight 3D convolution, sub-volume clipping, and multi-stage inference strategies to reduce resource consumption while maintaining high accuracy, making 3D structures more feasible in practical applications [2].

3.4. Detection - segmentation two-stage process

In addition to directly conducting end-to-end segmentation, some studies adopt a two-stage process of "detection - segmentation". This type of method first locates suspected nodules through the detection network, and then uses U-Net or its variants to conduct fine segmentation of these candidate regions [5]. Experiments show that compared with the single-stage segmentation process, the two-stage method has more advantages in the detection rate of micronodules and segmentation stability, especially in cases with a large number of lesions or complex chest structures, which is more conducive to reducing missed detections and improving overall sensitivity [5].

3.5. Convolutional - transformer hybrid structure and engineering improvements

In recent years, the hybrid structure of convolution and Transformer has gradually attracted attention. Taking S3TU-Net as an example, this network combines structural convolution, Transformer, and superpixel information to achieve results of 89.04% DSC, 90.70% mIoU, and 93.70% sensitivity on the DCC-IDRI dataset. Compared with multiple existing methods, it has increased by approximately 4.52% and 3.16% respectively, and the improvement is particularly significant in small-volume and complex-background cases [6].

At the engineering practice level, some improvements have focused on feature fusion and attention design. For instance, dynamic convolution, SimAM attention, AFF modules, and SKAttention structures have been introduced into the U-Net, enabling the network to respond more precisely to small targets and boundary regions. On the LUNA16 dataset, such improvements led to approximately a 1.46% increase in Dice, a 2.64% increase in mIoU, an 8.25% increase in recall rate, and a 1.31% increase in accuracy [9]. These results indicate that targeted structural modification and information fusion design based on U-Net can significantly improve the segmentation performance of pulmonary micronodules without greatly increasing the complexity of the model.

4. Exploration and discovery

Although the U-Net series models have made considerable progress in the segmentation task of pulmonary micronodules, from the current research and practical application situations, there are still several notable issues.

First of all, the imaging features of the small nodules themselves are not prominent. They usually have low contrast, blurred boundaries, and often adhere to small blood vessels, making it difficult for the model to accurately distinguish these weak feature regions [2]. Although the introduction of multi-scale structures, dense jump connections, and attention mechanisms can alleviate these problems to a certain extent, the space for improvement is still limited under the premise of insufficient original signals [3,4].

Secondly, the generalization problem caused by differences in data sources still exists. The differences in gray level and texture caused by different scanning parameters, devices, and imaging conditions can lead to significant performance degradation of the model across centers or datasets. This "domain shift" phenomenon has been observed in multiple works [2,5].

Thirdly, the complex background environment remains a difficulty. For instance, ground-glass nodules are often adherent to blood vessels, located near the pleura, and similar in gray scale to the surrounding tissues. It is extremely difficult to distinguish them merely from local textures, which can easily lead to undersegmentation or missegmentation [2,4]. Although the attention mechanism or the convolution-Transformer hybrid structure has improved in some samples, it is still difficult to completely avoid bias in typical complex cases [4,6].

Furthermore, the interpretability of the existing models still has deficiencies. Although teacher-student knowledge distillation helps improve model stability, the feature extraction process within deep learning is still difficult to present directly to doctors, which may affect the transparency of decision-making [7].

In recent years, research has gradually expanded from the optimization of a single structure to the design of more complete technical routes. For instance, integrate detection, segmentation, and classification into a unified framework to reduce information loss; Or, in cases with fewer annotations, the robustness of the model can be improved by combining semi-supervised and knowledge distillation methods [5,7]. Meanwhile, the accumulation of multi-center data and the standardization of imaging processes are regarded as the key foundation for further enhancing the stability of the model [2,6,7].

Overall, the U-Net series of methods has demonstrated its feasibility in the automatic segmentation of pulmonary micronodules, but there is still room for improvement in terms of stability, cross-domain generalization ability, and interpretability [10]. With the continuous development of network structures, data construction strategies, and learning paradigms, these methods are expected to be better integrated into clinical screening and follow-up work in the future [2,5,7].

5. Conclusion

This paper systematically reviews the main methods and current development status of automatic segmentation research on pulmonary micronodules, focuses on analyzing the application performance of U-Net and its derivative structures in different scenarios, and summarizes the current main technical paths from the perspectives of multi-scale feature fusion, attention mechanism, three-dimensional convolution design, and convolution-Transformer hybrid structure. Overall, in recent years, significant progress has been made in small target representation, boundary refinement, suppression of complex background interference, and cross-level feature integration, which has enabled deep learning methods to demonstrate increasingly high stability and practical value in the automatic recognition and segmentation of pulmonary micronodules.

Although the existing methods are constantly being improved, there are still some common challenges. For instance, the weak signal of tiny nodules themselves and their tight adhesion to blood vessels may still lead to blurred edges or structural confusion in the model. The domain differences brought about by different devices and imaging protocols make the cross-center generalization ability of the model uncertain. The internal representation process of deep networks is opaque, which also limits their interpretability and traceability in clinical Settings. These issues indicate that although the existing methods have performed well, there is still room for improvement before they can be fully and stably applied in multi-scenario clinical processes.

Future research can focus on more robust multi-scale structures, more efficient lightweight 3D modeling strategies, and multi-modal fusion mechanisms. Meanwhile, the accumulation of multi-center data, the establishment of standardized image acquisition processes, and explainable technologies that are more easily understood in clinical practice will also become important directions for promoting the practical implementation of the model. Overall, the automatic segmentation technology of pulmonary micronodules has broad application prospects in intelligent diagnosis, screening follow-up, and quantitative assessment. With the continuous maturation of the method system, its value in clinical practice is expected to be further enhanced.

References

[1] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) (pp. 234–241).

[2] Shen, Q., Zhang, X., Li, W., et al. (2023). Application progress of U-Net in pulmonary nodule segmentation. Computer Applications, 43(S1), 250–257.

[3] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2019). UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE Transactions on Medical Imaging, 38(8), 1856–1867.

[4] Hou, Y., et al. (2022). Pulmonary nodule image segmentation based on U-Net incorporating dual attention modules. Computer Science and Applications.

[5] Xie, H., Yang, D., Sun, N., et al. (2019). Automated pulmonary nodule detection in CT images using deep convolutional neural networks. Pattern Recognition, 85, 109–119.

[6] Wu, Y., Liu, X., Shi, Y., Chen, X., Wang, Z., Xu, Y., & Wang, S. (2025). S 3 TU-Net: Structured convolution and superpixel transformer for lung nodule segmentation. Medical & Biological Engineering & Computing, 63(12), 3777-3791.

[7] Liu, W., Zhang, L., Zhang, X., et al. (2025). A semi-supervised knowledge distillation model for lung nodule segmentation. Scientific Reports, 15, 562.

[8] Oktay, O. (2018). Attention U-Net: Learning where to look for the pancreas. arXiv preprint arXiv: 1804.03999.

[9] Li, X. (2025). Research on lung nodule image detection and segmentation algorithm based on deep learning (Master’s thesis).

[10] Wang, Y., Li, X., Chen, Z., et al. (2024). Enhanced U-Net with multi-attention fusion for precise segmentation of pulmonary micronodules in low-dose CT images. Medical Image Analysis, 92, 102115.