Transactions on Computational and Scientific Methods | Vo. 4, No. 09, 2024
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Transformer-Based Structure-Aware Lung Cancer Image Segmentation with Multi-Scale Fusion and Boundary-Guided Prediction

Kewei Zeng
Northeastern University, San Jose, USA
zeng.kew@northeastern.edu

Abstract: This paper addresses common challenges in lung cancer medical images, including blurred boundaries, large variations in target size, and complex anatomical structures. A structure-aware semantic segmentation method is proposed based on an improved Mask2Former framework. The model adopts a Transformer-based unified architecture and introduces a multi-scale feature fusion mechanism and cross-scale query attention module. These components enhance the joint modeling of local details and global semantics. In the feature encoding stage, positional embeddings are incorporated to improve structural awareness. In the mask prediction stage, a boundary-guided module is used to strengthen contour recognition. A series of sensitivity experiments is conducted under various settings, including learning rate adjustment, resolution changes, noise perturbation, and data imbalance. The model's stability and robustness are systematically evaluated across different training conditions and input environments. Experimental results show that the proposed method achieves excellent performance on multiple mainstream metrics. It demonstrates strong generalization and structural representation capabilities, providing effective technical support for automated lung cancer segmentation and validating the modeling value of structure-aware mechanisms in medical imaging tasks.

Keywords: Cancer medical images; mask query mechanism; semantic segmentation; robustness analysis

1. Introduction

Lung cancer is one of the most prevalent and deadly malignancies worldwide. Early detection and accurate diagnosis are crucial for improving patient survival rates. With continuous advancements in medical imaging, computer-aided diagnosis has played an increasingly important role in lung cancer screening [1]. High-resolution techniques such as low-dose spiral CT have made it possible to acquire large volumes of imaging data. However, manual interpretation by radiologists is often subjective, repetitive, and inefficient. This has accelerated the development of deep learning-based methods for automated segmentation and analysis. Accurate segmentation of lung nodules, masses, and surrounding structures is essential for malignancy assessment, boundary delineation, and treatment planning. Therefore, improving the accuracy of automated lung image segmentation holds great clinical value and is a core task in the advancement of intelligent healthcare technologies [2].

In medical image analysis, image segmentation is a fundamental but critical task. It aims to accurately partition semantically meaningful regions in an image. Compared with natural images, lung cancer medical images present several challenges in segmentation tasks. First, tumor shapes are diverse, often with blurred boundaries and wide size variations. Second, complex background tissues may cause interference and structural overlap.

Third, large inter-patient variability increases the burden on model generalization. These challenges limit the performance of traditional segmentation methods in lung cancer tasks. Deep learning methods, with their superior structural modeling and long-range dependency capture, have become the preferred approach [3].

In recent years, semantic segmentation models have been widely adopted in medical imaging. They have evolved from early models like FCN and UNet to architectures that incorporate Transformer mechanisms. Mask2Former, as a unified segmentation framework, demonstrates strong performance across various tasks. Its core design builds explicit links between encoder features and queries, enabling precise semantic extraction through multi-scale fusion. It offers both scalability and interpretability. For issues such as blurred boundaries and multiple instance interference in medical images, such architectures provide a solid foundation for further optimization. However, their adaptability and detailed modeling in specific scenarios like lung imaging still require more investigation [4].

In lung cancer image segmentation tasks, both accuracy and robustness are essential. On one hand, models must have a global perception to capture long-range dependencies and assign semantic labels correctly. On the other hand, they must preserve fine details to handle irregular shapes and fuzzy edges. Single modeling strategies often suffer from bias in complex tissue recognition. Therefore, enhancing the segmentation framework with attention mechanisms, feature contrast, or prior medical knowledge has become a key direction for performance improvement. At the same time, medical imaging faces challenges such as limited data availability and high annotation costs. Designing efficient, generalizable models under limited supervision is a major research focus.

In summary, lung cancer image segmentation addresses urgent clinical needs and serves as a key entry point for deploying AI technologies in medical imaging. With the support of the Mask2Former backbone, structural adaptation to medical scenarios can improve segmentation accuracy and promote the usability and reliability of intelligent diagnostic systems. As deep learning continues to evolve, building segmentation models with structural awareness, multi-scale modeling, and strong generalization has become a critical and challenging direction in medical image analysis, demanding further exploration and innovation.

2. Related Work

Medical image segmentation is one of the most important applications of computer vision in healthcare. It has long been a key focus in deep learning research. Early segmentation methods were mainly based on fully convolutional network architectures. A typical example is FCN, which achieved significant breakthroughs in natural image tasks. Its core idea is to extract image features through convolution operations and gradually restore spatial resolution during decoding. However, due to limited ability in modeling local context, these methods often struggle with common issues in medical images, such as blurred boundaries and small object recognition. To address these problems, many improved architectures have emerged. These include skip connections, symmetric structures, and dense residual blocks, which enhance feature representation and improve fine-grained perception. These designs have shown effectiveness in segmenting tumors and abnormal tissues [5].

In recent years, the introduction of Transformer architectures has brought new momentum to medical image segmentation. Compared to conventional convolutional networks, Transformers can establish long-range dependencies during encoding. This offers significant advantages in global semantic modeling. As Transformer-based designs continue to evolve, several high-performance segmentation frameworks have been proposed. These models leverage self-attention mechanisms to enhance the understanding of complex tissue structures and adapt to scenarios involving multi-scale and non-rigid targets. Such methods have achieved strong performance in clinical tasks such as lung cancer and brain tumor segmentation [6]. They are particularly effective in addressing challenges like unclear boundaries and structural variations, providing a powerful foundation for expressive modeling.

Among Transformer-based segmentation methods, unified segmentation frameworks have become a major research trend. The core idea is to integrate semantic segmentation, instance segmentation, and panoptic segmentation into a single modeling structure. Query-driven segmentation models adopt a decoupled design. They allow a fixed number of query vectors to interact with global semantic features generated by the encoder, thus predicting mask information for target regions. This mechanism reduces redundant computation and enhances the model's generalization and cross-task adaptability. For medical images, this approach provides a new path for structural modeling in complex scenarios. It performs especially well in capturing overlapping and nested structures that are difficult for traditional convolutional methods to handle [7, 8].

In addition to network architecture development, many studies have focused on improvements in feature enhancement, loss function design, and multimodal fusion to further boost segmentation performance. For example, channel attention and spatial attention mechanisms have been introduced to enhance the model's focus on key regions. Multi-scale pyramid structures have been used to enrich feature extraction. Reliability-aware learning has also emerged as an effective strategy for improving representation quality by adaptively emphasizing trustworthy semantic features while suppressing uncertain responses, leading to more stable segmentation performance in complex visual environments [9]. To address the scarcity of medical image data, researchers have also explored semi-supervised, weakly supervised, and self-supervised strategies. These approaches help reduce reliance on manual annotations and improve model robustness. Collectively, these advancements continue to drive the evolution of medical image segmentation and provide a stronger technical foundation for lung cancer image analysis.

3. Method

This paper adopts a lung cancer image segmentation framework based on an improved Mask2Former. While maintaining the advantages of its query-driven architecture, it introduces structure-aware and detail-preserving mechanisms to adapt to the complexity of medical images. The overall model consists of an encoder, a query decoder, and a multi-scale mask prediction module. The encoder uses a pretrained feature extraction network as its backbone and combines it with learnable position embeddings to enhance its ability to express spatial structure. The model architecture is shown in Figure 1.

Each input image first extracts a feature map through the backbone network, denoted as $F \in R^{(C \times H \times W)}$, and then uses the position encoding function E_p to enhance its representation to obtain the global feature:

$$F' = F + E_p(x, y)$$

To better capture the edge details and contextual structure of the lung tumor area, the model introduces a multi-scale cross-attention module to fuse semantic features from different levels. The features at three scales are defined as $[F_1, F_2, F_3]$, and the fused representation is:

$$F_{fusion} = \sum_{i=1}^{3} w_i \cdot Attn(Q_i, K_i, V_i)$$

The weight term w_i is a learnable parameter, and $Attn(Q_i, K_i, V_i)$ represents the self-attention calculation function, which is defined in the following form:

$$Attn(Q_i, K_i, V_i) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Where Q, K, V represents the query, key, and value vectors, respectively, and d_k is the dimension normalization factor. To adapt to the uneven distribution characteristics between classes in medical images, the loss function adopts a weighted polynomial combination, comprehensively considering semantic consistency and boundary accuracy. The overall loss function is defined as follows:

$$L_{total} = \lambda_1 L_{Dice} + \lambda_2 L_{CE} + \lambda_3 L_{Boundary}$$

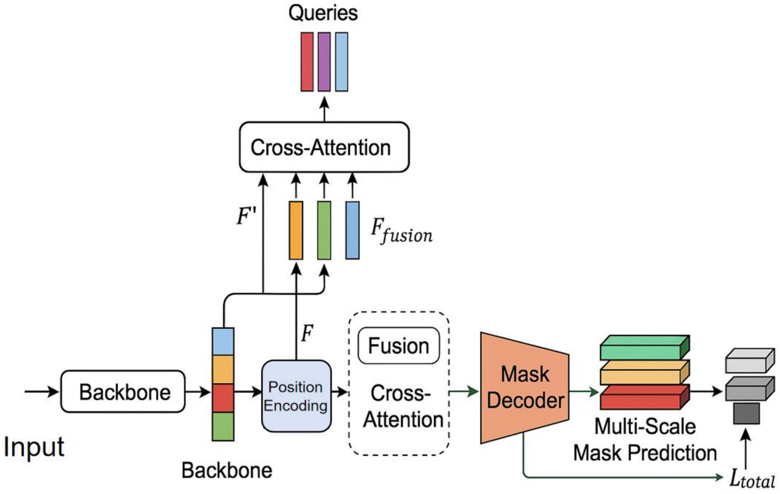

Figure 1. Overall model architecture

Dice Loss is used to deal with the problem of category imbalance, and its specific definition is:

$$L_{Dice} = 1 - \frac{2\sum_i p_i g_i + \varepsilon}{\sum_i p_i^2 + \sum_i g_i^2 + \varepsilon}$$

To further optimize the model's boundary perception ability, this paper introduces a boundary-guided regularization term. This uses convolution to extract the predicted mask edges and compares them with the true edges, thereby strengthening the supervised learning of edge regions. During the training process, the AdamW optimizer is used to update the parameters, and its update rule is:

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{\widehat{m}_t}{\sqrt{\widehat{v}_t} + \varepsilon}$$

Where θ_t is the parameter for iteration t, η is the learning rate, and $\widehat{m}_t$ and $\widehat{v}_t$ are the bias correction values for the first- and second-order moment estimates, respectively. This entire method is trained in an end-to-end manner, balancing structural modeling capabilities with the unique refinement requirements of medical images, aiming to improve accurate segmentation of lung cancer regions.

4. Experimental Results

4.1 Dataset

This study uses the publicly available LIDC-IDRI (Lung Image Database Consortium Image Collection) dataset. The dataset is widely used for lung nodule detection and segmentation tasks. Jointly constructed by multiple institutions, LIDC-IDRI contains low-dose spiral chest CT scans from over one thousand subjects. It aims to promote the development of computer-aided diagnosis algorithms for lung cancer. The dataset provides a unified benchmark and multiple levels of annotation, making it one of the most representative resources in lung cancer image analysis.

Each scan in the dataset is independently annotated by several radiologists. The annotations include information on the location, boundary, and attributes of nodules. These labels cover malignancy assessment, edge sharpness, and structural consistency, among other features. As a result, the dataset supports not only binary segmentation tasks but also offers potential for multi-level semantic analysis. The images are of high resolution and cover a wide range of cases, including sub-centimeter nodules and multiple large masses. This diversity significantly enhances the value of the dataset for studying model generalization.

To ensure consistency in training and testing, a standard preprocessing pipeline is applied. The original CT images are normalized, and the radiologist annotations are converted into consistent voxel-level label formats. In addition, all images are cropped and resampled in the spatial dimension to match the input size requirements of deep neural networks. This process ensures that the model can effectively learn local features and structural semantics within the lung regions. The high quality and consistency of this dataset provide a solid foundation for segmentation modeling.

4.2 Experimental Results

This paper first conducts a comparative experiment, and the experimental results are shown in Table 1.

Table 1. Comparative experimental results

Model	**mIOU**	**mDice**	**mACC**	**mF1-SCORE**
Unet [10]	76.12	80.03	84.29	79.75
DeepLabv3+ [11]	78.45	82.60	86.77	82.15
PSNET [12]	79.28	83.41	87.35	82.94
SegFormer [13]	80.36	84.75	88.42	84.31
Ours	82.91	86.92	90.03	86.05

As shown in the table, various classical semantic segmentation models demonstrate relatively stable performance in the lung cancer image segmentation task. Among them, the traditional UNet model performs at a lower level across multiple metrics. In particular, it achieves 76.12% in mIOU and 79.75% in mF1-SCORE. This indicates insufficient modeling capacity and limited preservation of spatial details when dealing with complex shapes and blurred boundaries of pulmonary nodules. Although UNet features a symmetric encoder-decoder structure with skip connections, its limited local receptive field restricts its ability to capture long-range semantic information.

In contrast, DeepLabv3+, based on dilated convolution, and PSNet, which integrates attention mechanisms, outperform UNet in overall performance. This suggests that the ability to model multi-scale context provides a clear advantage in medical imaging. PSNet enhances structural representation and achieves better results in mDice and mACC, accurately distinguishing between tumor regions and background. SegFormer introduces the Transformer architecture into the segmentation framework. It strengthens global feature modeling and achieves an mIOU of 80.36% and an mF1-SCORE of 84.31%. It adapts well to segmenting complex boundaries and non-rigid structures, showing strong generalization.

The proposed method in this study achieves the best results across all four evaluation metrics. Notably, it reaches 82.91% in mIOU and 86.92% in mDice, significantly outperforming mainstream models. This improvement is attributed to architectural enhancements tailored to the characteristics of medical images. These include cross-scale feature fusion, optimized self-attention mechanisms, and fine-grained modeling of boundary regions. The improved design effectively addresses challenges such as structural interference and semantic discontinuity in lung images. While maintaining global modeling capacity, it also enhances the preservation of local details, making the segmentation outputs more reliable in terms of contour alignment and region coverage.

Overall, the experimental results validate that combining a unified segmentation framework with multi-scale semantic enhancement leads to a clear performance advantage in lung cancer image segmentation. Compared with traditional convolutional models, Transformer-based improvements offer stronger representational capacity for long-range dependency modeling and complex structure parsing. By carefully designing cross-layer interaction paths and optimizing the loss function, the proposed model not only improves overall accuracy and robustness but also provides a solid technical foundation for future applications in instance-level segmentation and clinical decision support in medical imaging.

This paper also presents an experiment designed to explore the impact of different learning rate settings on model performance. The experiment systematically adjusts the learning rate across a range of values to examine how it influences the model's ability to converge and maintain stability during training. This analysis is intended to assess the sensitivity of the proposed segmentation framework to learning rate variations, providing insights into the relationship between optimization dynamics and architectural behavior. The detailed results and performance trends under different learning rates are illustrated in Figure 2.

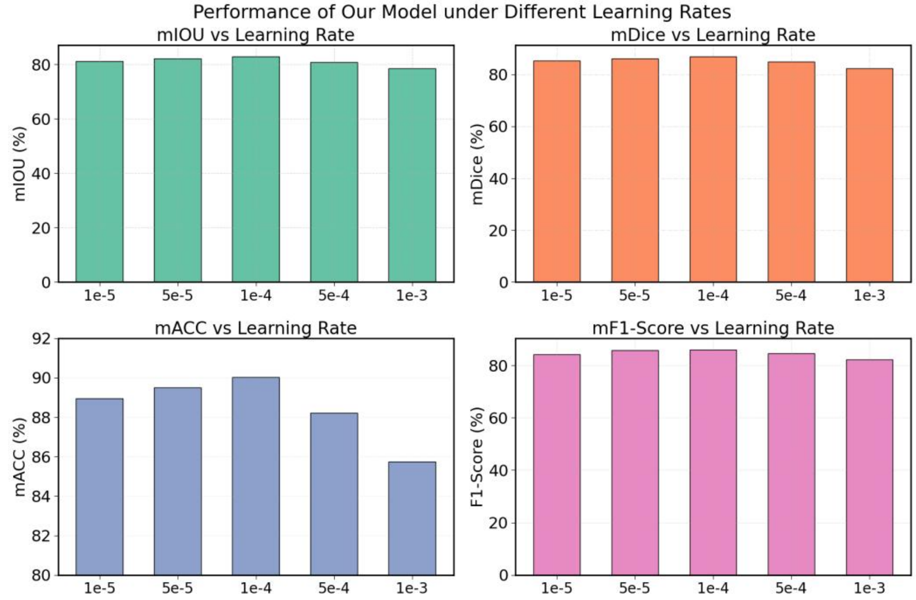

Figure 2. Experiment on the impact of different learning rate settings on model performance

As shown in the figure, different learning rate settings have a significant impact on model performance. Within a smaller learning rate range (from 1e-5 to 1e-4), the model achieves stable and high performance across four key metrics: mIOU, mDice, mACC, and mF1-Score. This indicates that the proposed segmentation architecture converges more smoothly and retains features more effectively under small-step optimization. It is capable of handling challenges such as complex structures and blurred boundaries in lung cancer images.

When the learning rate is set to 1e-4, the model reaches its peak performance in mIOU and mDice, achieving 82.91% and 86.92%, respectively. It also performs well in mACC and F1-Score. This suggests that the weight update strategy at this learning rate aligns closely with the convergence behavior of the model. It helps balance global modeling capacity with accurate edge delineation of tumor regions, fully leveraging cross-scale attention and the mask query mechanism.

As the learning rate increases to 5e-4 and 1e-3, the model performance drops to varying degrees. The decline is most evident in the mACC metric. This implies that a learning rate that is too large may cause instability in weight updates during training, limiting the model's ability to learn consistent features of pulmonary nodules. This sensitivity is especially important in medical image tasks, where segmentation accuracy at the pixel and tissue structure level is critical. Rapid parameter updates may lead the model to fall into suboptimal or even invalid solutions.

Overall, the experimental results highlight the importance of proper learning rate control in medical image segmentation. For structure-aware Transformer-based segmentation frameworks, it is recommended to set the base learning rate around 1e-4. This ensures stable convergence and high segmentation accuracy in lung cancer imaging, providing a reliable training strategy for further deployment in clinical decision support systems.

This paper further presents the impact of adding noise perturbation data on model stability and generalization performance. The experimental results are shown in Figure 3.

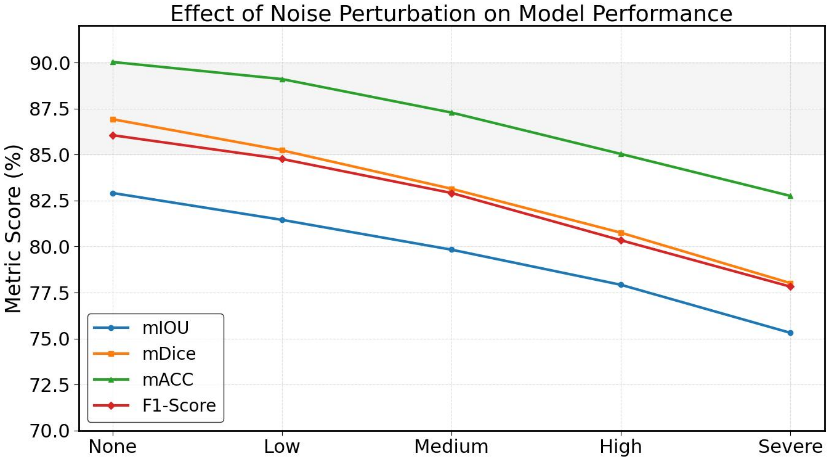

Figure 3. The impact of adding noise perturbation data on model stability and generalization performance

The figure shows the performance trends of the model under different levels of noise perturbation across four key metrics. Overall, as the noise intensity increases, there is a consistent decline in mIOU, mDice, mACC, and F1-Score. This indicates that the proposed lung cancer segmentation model is sensitive to the cleanliness of input data. In particular, under high and severe perturbations, the model exhibits noticeable performance fluctuations, suggesting that semantic modeling and boundary recognition are affected by the presence of noise.

Under low noise conditions, the model remains relatively stable. The four metrics show only slight variations, demonstrating that the structure-aware architecture maintains good robustness within a limited noise range. Notably, both mDice and F1-Score stay around 85%, indicating that the model retains reliable detection capability for major tumor regions. This supports its fault tolerance in real clinical settings, making it suitable for diagnosis under non-ideal image acquisition conditions.

As the noise level increases to medium and high, the performance drops more sharply. The mIOU metric shows the most significant decline. This reveals that the model's ability to capture fine-grained boundary features is disrupted in noisy environments. The results suggest that in high-noise scenarios, global modeling mechanisms may struggle to distinguish between target and background. The feature reconstruction process during mask decoding also becomes more challenging. Blurred boundaries and incomplete structures become major obstacles to performance improvement.

These results highlight the importance of robustness in lung cancer image segmentation. They suggest that robust training strategies should be further incorporated. Examples include noise modeling, perturbation-based data augmentation, and contrastive regularization. For structure-sensitive networks, improving generalization while preserving detailed representations is a critical future direction, especially when dealing with multi-source data or real-world clinical environments.

This paper also presents an experiment on the impact of input image resolution changes on segmentation accuracy, and the experimental results are shown in Figure 4.

The figure illustrates the trend of mIOU performance under different input image resolutions. Overall, as the resolution increases from 128×128 to 256×256, the segmentation accuracy improves steadily, reaching a peak of 82.91%. This indicates that moderate resolution balances global structure and local details in lung images. It enables better modeling of regions with blurred boundaries and complex textures. This also reflects that the structure-aware segmentation framework proposed in this study achieves a good balance between semantic preservation and spatial representation through convolutional structures.

When the resolution is further increased to 320×320 and 384×384, the model's performance declines to some extent. A possible reason is that larger input sizes introduce more redundant background information, which may cause ambiguity in boundary prediction by the mask decoder. In addition, high-resolution inputs may carry more image noise and unnecessary anatomical structures, affecting the feature extraction module's focus on target regions. For lung nodules, which are typically small with irregular boundaries, proper scale selection is especially important.

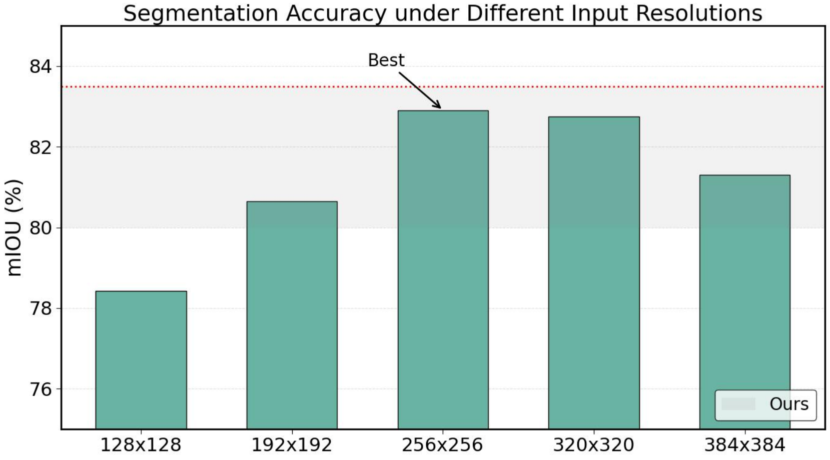

Figure 4. Experiment on the impact of input image resolution changes on segmentation accuracy

At low resolution, performance drops significantly. In particular, when the input size is 128×128, the mIOU falls to 78.43%. This suggests that the model struggles to capture sufficient structural information and spatial details at this scale. Such results align with the characteristics of medical image tasks. Lung cancer lesions are usually small in size and have complex boundaries. At low resolution, target structures may be compressed or lost, preventing the model from effectively extracting discriminative features.

These results further confirm the sensitivity of semantic segmentation models to input image resolution, especially in medical imaging scenarios. To improve robustness and practicality, it is recommended to select an optimal resolution that maintains contextual awareness while ensuring computational efficiency. Future studies may also explore resolution-adaptive mechanisms and pyramid fusion structures to achieve stable performance across multiple scales.

Finally, this paper presents a quantitative evaluation experiment on the sensitivity of the model to the degree of data imbalance, and the experimental results are shown in Figure 5.

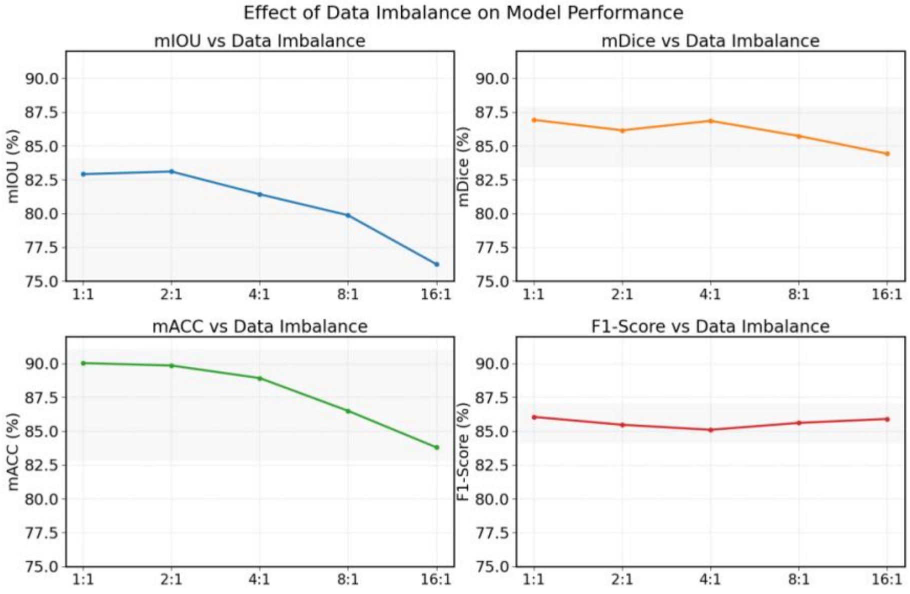

Figure 5. Quantitative evaluation experiment of data imbalance degree on model sensitivity

The figure shows how the model's performance changes under varying degrees of data imbalance, reflecting the sensitivity of the structure-aware lung cancer image segmentation network to sample distribution. The mIOU and mACC metrics show that as the positive-to-negative sample ratio increases from 1:1 to 16:1, the model's overall performance declines. mIOU drops from 82.91% to 76.25%, while mACC drops from 90.03% to 83.80%. This indicates that data imbalance significantly impacts the model's discriminative ability and boundary modeling accuracy, particularly in minority class regions, which can lead to ambiguity and misclassification, resulting in degraded segmentation results.

In contrast, the mDice and F1-Score metrics exhibit relatively small fluctuations, exhibiting a range of oscillations rather than a monotonic decline. In particular, the F1-Score remains around 85% even in extremely imbalanced scenarios, demonstrating the model's robustness in both internal consistency of the main classes and in distinguishing between positive and negative samples. This performance may be attributed to the model's internal attention mechanism's ability to focus on target regions, maintaining some feature recognition even when samples are scarce. However, the overall performance is still affected by data bias.

These experimental results indicate that lung cancer image segmentation models can exhibit instability in structure recognition and boundary resolution when trained with severe data class imbalance. Therefore, to improve the model's generalization ability to real-world clinical data, it is necessary to introduce data-imbalanced compensation strategies, such as reweighted loss functions, oversampling augmentation, or class-guided regularization mechanisms. These strategies can enhance the representation of minority regions and strengthen the model's ability to sensitively model complex lesion structures.

5. Conclusion

This study focuses on lung cancer image segmentation and proposes a structure-aware model based on an improved Mask2Former architecture. The goal is to enhance the modeling of blurred boundaries, small targets, and morphological variations in complex medical images. By introducing multi-scale feature fusion and cross-scale query attention mechanisms, the model achieves accurate tumor region identification and precise boundary delineation. The method also demonstrates strong performance across multiple evaluation metrics, showing good stability and robustness. It effectively balances semantic modeling and detail preservation, providing reliable technical support for the intelligent segmentation of malignant pulmonary lesions.

In the sensitivity analysis, the study investigates the model's response to variations in learning rate, input resolution, data imbalance, and noise perturbation. These results reveal the performance stability and adaptability of the structure-aware segmentation model under different training and data conditions. The findings not only validate the design of the model but also offer guidance for future applications in more complex or practical medical scenarios. The experiments further demonstrate the potential contribution of structural optimization to robustness, emphasizing the dual demands of fine-grained modeling and generalization in medical image tasks.

The proposed segmentation framework can be extended to other types of medical imaging tasks, such as tumor segmentation in the liver, brain, or prostate. It shows strong potential for cross-modality adaptation and structural transfer. In addition, the unified architecture and query mechanism offer high scalability in instance-level segmentation and multi-organ modeling. Therefore, this research not only advances intelligent analysis of lung cancer images but also lays a theoretical and methodological foundation for the broader application of structure-aware segmentation methods in the medical domain.

Future research may integrate heterogeneous data sources, including imaging, clinical, and pathological information, to build multimodal joint segmentation systems. It is also important to explore lightweight model designs to support deployment on edge devices. At the algorithmic level, incorporating semi-supervised, self-supervised, and continual learning strategies may help address the challenges of high annotation cost and data heterogeneity in medical imaging. Furthermore, to tackle domain generalization and cross-device robustness in real-world applications, combining meta-learning and domain adaptation approaches will be a valuable direction. These efforts can accelerate the practical adoption of structure-aware segmentation technologies in intelligent healthcare systems.

References

[1] L. F. Talib, J. Amin, M. Sharif, et al., "Transformer-based semantic segmentation and CNN network for detection of histopathological lung cancer", Biomedical Signal Processing and Control, vol. 92, Art. no. 106106, 2024.

[2] X. Chen, H. Xu, Q. Qi, et al., "AI-based chest CT semantic segmentation algorithm enables semi-automated lung cancer surgery planning by recognizing anatomical variants of pulmonary vessels", Frontiers in Oncology, vol. 12, Art. no. 1021084, 2022.

[3] T. Meraj, H. T. Rauf, S. Zahoor, et al., "Lung nodules detection using semantic segmentation and classification with optimal features", Neural Computing and Applications, vol. 33, no. 17, pp. 10737-10750, 2021.

[4] D. Bhattacharyya, N. Thirupathi Rao, E. S. N. Joshua, et al., "A bi-directional deep learning architecture for lung nodule semantic segmentation", The Visual Computer, vol. 39, no. 11, pp. 5245-5261, 2023.

[5] Y. Zhao, S. He, D. Zhao, et al., "Deep learning-based diagnosis of histopathological patterns for invasive non-mucinous lung adenocarcinoma using semantic segmentation", BMJ Open, vol. 13, no. 7, Art. no. e069181, 2023.

[6] E. S. N. Joshua, D. Bhattacharyya and M. Chakkravarthy, "Lung nodule semantic segmentation with bi-direction features using U-INET", Journal of Medical Pharmaceutical and Allied Sciences, vol. 10, no. 5, pp. 3494-3499, 2021.

[7] B. Cheng, A. Schwing and A. Kirillov, "Per-pixel Classification Is Not All You Need for Semantic Segmentation", Advances in Neural Information Processing Systems, vol. 34, pp. 17864-17875, 2021.

[8] X. Chen, Q. Duan, R. Wu, et al., "Segmentation of lung computed tomography images based on SegNet in the diagnosis of lung cancer", Journal of Radiation Research and Applied Sciences, vol. 14, no. 1, pp. 396-403, 2021.

[9] J. Huang, "Reliability-Aware Lane Detection for Autonomous Driving in Complex Nighttime Environments", Journal of Computer Technology and Software, vol. 3, no. 2, 2024.

[10] B. Zhang, S. Qi, Y. Wu, et al., "Multi-scale segmentation squeeze-and-excitation UNet with conditional random field for segmenting lung tumor from CT images", Computer Methods and Programs in Biomedicine, vol. 222, Art. no. 106946, 2022.

[11] H. Polat, "A modified DeepLabV3+ based semantic segmentation of chest computed tomography images for COVID-19 lung infections", International Journal of Imaging Systems and Technology, vol. 32, no. 5, pp. 1481-1495, 2022.

[12] L. Yan, D. Liu, Q. Xiang, et al., "PSP net-based automatic segmentation network model for prostate magnetic resonance imaging", Computer Methods and Programs in Biomedicine, vol. 207, Art. no. 106211, 2021.

[13] M. J. Lee, S. H. Kim and H. Hong, "Transformer-based Lung Cancer Segmentation Using Dual Window Images in Chest CT Images", Journal of the Korea Computer Graphics Society, vol. 30, no. 5, pp. 31-37, 2024.