

EPIDSeg-Net: A Multi-Modal Fusion Framework Based on DRR Guidance in Radiotherapy is Used for Precise Segmentation of MV-EPID Lung Targets

Qianjia Huang, MEng¹; Xinye Ni, MD²

¹Changzhou University, Changzhou, China; ²Nanjing Medical University, Nanjing, China

Abstract

This study proposes a multimodal segmentation framework, named EPIDSeg-Net, which integrates Digitally Reconstructed Radiograph (DRR) images of pulmonary tumors with Electronic Portal Imaging Device (EPID) images to assist target segmentation, and further compares morphological changes in segmented targets across different radiotherapy stages to achieve precise quantification of the radiotherapy target volume. Experimental results demonstrate that the framework effectively fuses multimodal information, achieving high-precision localization and clear delineation while preserving anatomical details. Quantitative evaluation yields excellent performance. The model excels in preserving gradient information, regional integrity, and fine details; it effectively suppresses feature loss and reduces missed segmentation rates, with notable improvements in both subjective and objective metrics.

Introduction

Primary lung cancer is a common respiratory malignancy, and radiotherapy is a key treatment. Post-treatment imaging evaluation, which quantifies tumor regression and residual volume while detecting resistance or recurrence early, is essential for guiding therapy and optimizing plans dynamically. In precision radiotherapy, EPID enables real-time monitoring but suffers from low contrast, leading to blurred tumor boundaries and unclear organ-at-risk contours. When tumors shift or change shape during fractionated treatment, single-modality EPID alone cannot reliably support target delineation, hampering adaptive planning and personalized decisions. To address this, multimodal fusion integrating high-resolution DRR images offers a powerful solution. DRR verifies setup and target coverage, and its superior soft-tissue contrast reveals biological tumor boundaries, providing vital anatomical references for EPID. Its clinical value lies in three aspects: (1) dynamic adaptation through DRR-EPID fusion to assess tumor regression; (2) correction of scatter noise for accurate dose reconstruction; and (3) enhanced segmentation via a dual-channel deep learning framework, enabling precise “image-dose-response” analysis and moving radiotherapy from geometry- to biology-guided practice.

Methods and Materials

Dataset: An in-house private DRR-EPID-Seg dataset was used, comprising 350 paired EPID and DRR images with a resolution of $1 \times 512 \times 512$. EPID images were acquired at 6 MV / 2 MU, and DRR images were generated from planning CT (120 kVp / 150 mAs). Preprocessing included anonymization, CT calibration, inter-modal histogram matching, and non-rigid registration.

Network Architecture: The proposed EPIDSeg-Net is based on a U-shaped Swin-Transformer and consists of an encoder, a multi-scale feature layer, and a decoder. The encoder adopts a dual-branch design: CNN for local texture extraction and Swin-Transformer for global semantic capture, with a Dual Attention Module (DAM) to adaptively adjust modality weights. The multi-scale layer incorporates Large-Kernel Grouped Attention Gating (LKG-Gate) to enhance local context and Multi-Path Feature Extraction (MPFE) to improve feature robustness via parallel structures. The decoder restores resolution through upsampling and skip connections, as **Figure 1**.

Training Details: Input resolution was reduced to 128×128 . The model was implemented with PyTorch on an A100 GPU, using the AdamW optimizer (lr=0.001), batch size of 1, and trained . Data augmentation included random rotation, cropping, and flipping, with five-fold cross-validation. The loss function was a composite of cross-entropy, Dice loss, and Focal Tversky loss with weights 0.5:0.3:0.2 and focal parameters q=0.3, n=0.7, m=2.

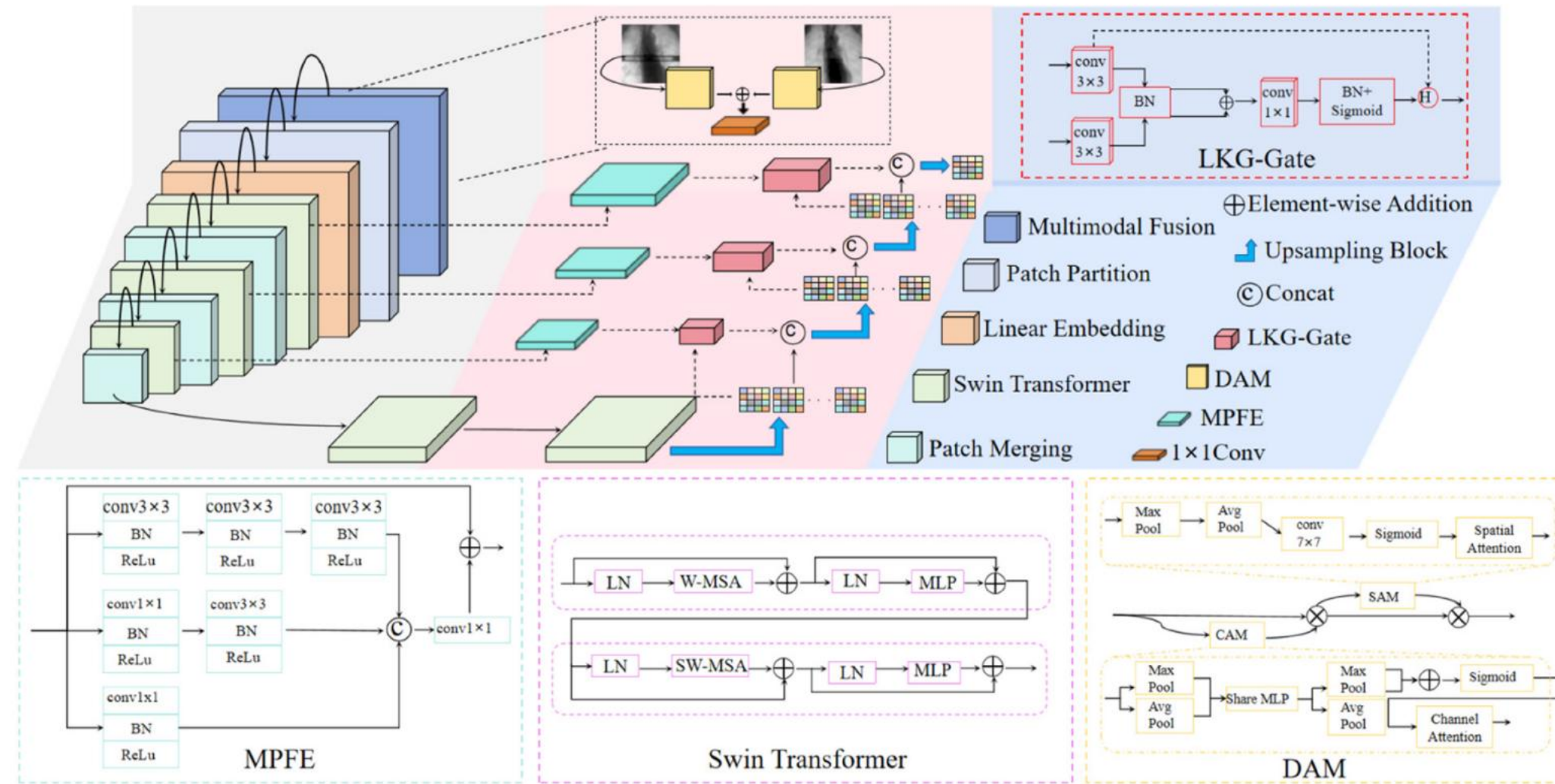


Figure 1. The architecture of proposed method.

Results

This study conducted a comprehensive evaluation on an in-house dataset containing 350 paired EPID and DRR images. Five-fold cross-validation demonstrated that the model achieved excellent average segmentation performance: DICE coefficient of 93.2%, IOU of 86.0%, HD95 of 9.4, cross-entropy loss of 0.352, and SENCE of 0.828. Compared with mainstream methods such as U-Net, U-Net++, U-Net3+, EANet, TransUnet, and Hiformer, our model achieved the best performance across all core metrics, with a particularly notable reduction in HD95, indicating more precise boundary delineation, as **Chart 1** and **Figure 2**. Ablation experiments further confirmed the independent contributions of the DAM, MPFE, LKG-Gate, and FTLoss modules; progressively adding each module yielded significant gains, with DICE improving from 91.2% to 93.2% and HD95 decreasing from 15.9 to 9.4, while the training process remained stable without overfitting. Overall, the results demonstrate that the model achieves state-of-the-art performance in accuracy, boundary quality, and generalization ability, meeting the requirements for precise clinical segmentation.

Methods	DICE(%)	HD95	IOU(%)	CE	SENSE
Unet ^[1]	88.6	19.2	80.3	0.652	0.513
Unet3+ ^[2]	89.3	17.9	81.2	0.599	0.542
Unet2+ ^[3]	92.2	13.6	84.5	0.489	0.795
EANet ^[4]	91.2	15.3	84.2	0.494	0.786
TransUnet ^[5]	92.6	13.4	84.9	0.472	0.803
Hiformer ^[6]	92.8	11.5	85.3	0.404	0.812
Ours	93.2	9.4	86.0	0.352	0.828

Chart 1. Comparative Results of Different Methods

Conclusions

This study proposes a novel multi-modality aided segmentation network, named EPIDSeg-Net, for segmenting treatment targets from low-quality EPID images. The model integrates the strengths of CNN and Swin Transformer architectures to extract local image features effectively. Furthermore, it incorporates specifically designed modules—DAM, MPFE, and LKG-Gate—that enable sequential feature transformation, achieving multi-modal feature fusion, multi-scale feature extraction, and large-field convolution optimization. These components collectively facilitate the modeling of long-range dependencies across modalities and the learning of multi-scale feature representations. Ablation studies confirm the effectiveness of each key component, and evaluations on the DRR-EPID-Seg dataset demonstrate superior performance over existing segmentation methods. Future work will focus on further optimizing the multi-modal segmentation framework to enhance its clinical applicability and reduce target segmentation response time.

Discussion

EPIDSeg-Net establishes a precise radiotherapy support system that integrates automated efficiency with clinical reliability to ensure segmentation outcomes meet clinical real-time decision-making requirements. However, this study was conducted under controlled single-center, single-vendor, and fixed-energy conditions. Caution should therefore be exercised when extrapolating these findings to other devices, beam energies, or more complex motion scenarios. Evaluating cross-vendor and cross-energy generalization capabilities, along with performance validation in complex motion scenarios, will be critical directions for subsequent multi-center or prospective EPIDSeg-Net research. Furthermore, systematic exploration of integrating EPIDSeg-Net with advanced motion management frameworks represents a meaningful future research direction. This prudent and well-defined approach—compared to potentially insufficient analyses under current data constraints—more effectively ensures the reliability and reproducibility of our conclusions while providing a clear development path for future work. Subsequent efforts will focus on incorporating multi-institutional data and lightweight model designs to reduce dependency on auxiliary modalities while maintaining accuracy, extending the framework to achieve high-precision target localization, and investigating optimized multi-modal fusion strategies.

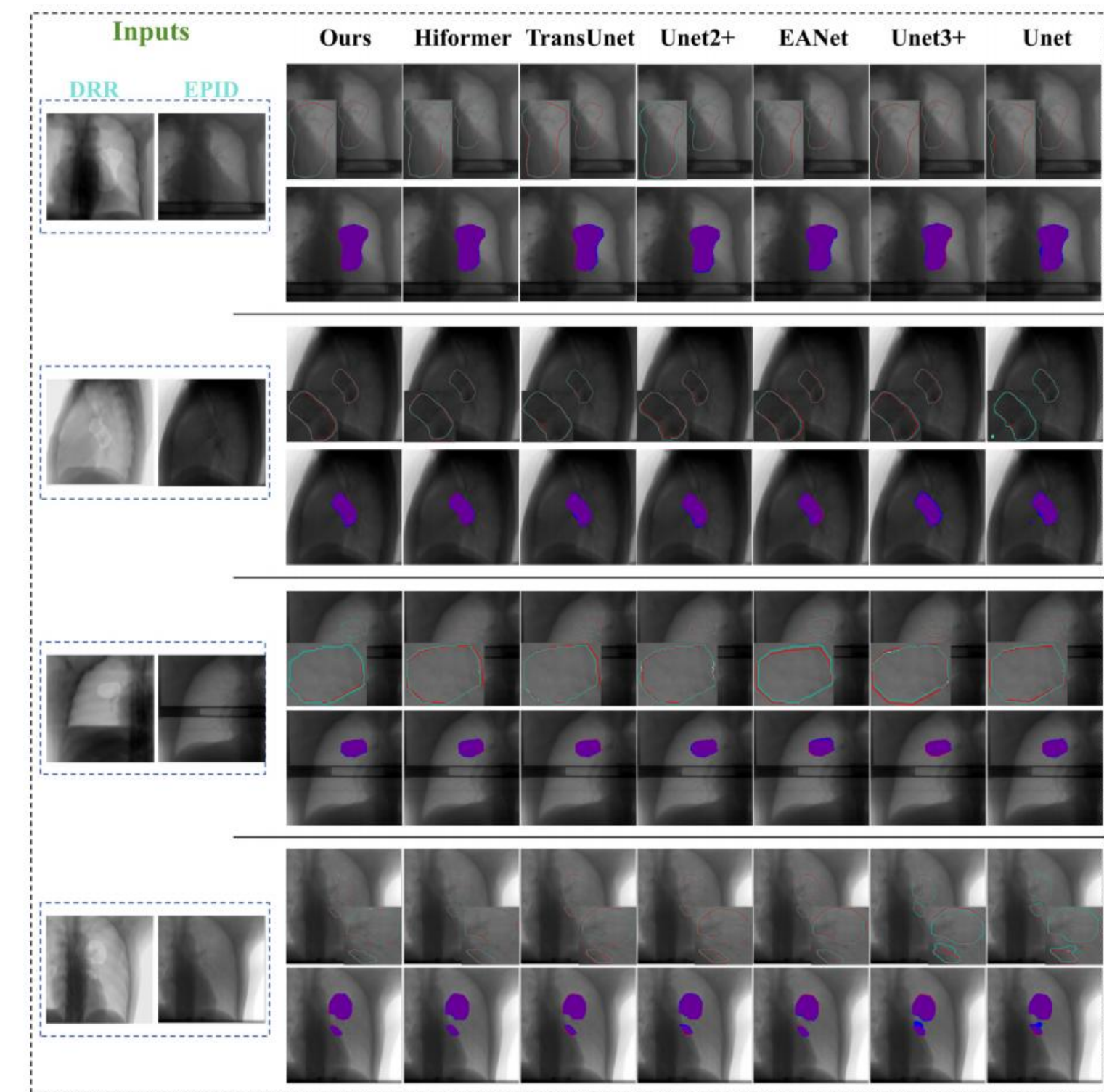


Figure 2. TSegmentation result comparison.

Contact

Qianjia Huang
Email: 424541674@qq.com

References

1. Ronneberger O, Fischer P, Brox T. U-net: convolutional net-works for biomedical image segmentation[C]. International Conference on Medical image computing and computer- assisted intervention. Cham: Springer international publishing; 2015: 234-241.
2. Huang H, Lin L, Tong R, et al. Unet 3+: a full-scale connected unet for medical image segmentation[C]. ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE; 2020: 1055-1059.
3. Zhou Z, Rahman Siddiquee M M, Tajbakhsh N, et al. Unet++: a nested u-net architecture for medical image segmentation[C]. International Workshop on Deep Learning in Medical Image Analysis. Springer International Publishing; 2018: 3-11.
4. Wang K, Zhang X, Zhang X, et al. EANet: iterative edge attention network for medical image segmentation[J]. Pattern Recognit. 2022;127:108636. doi:10.1016/j.patrec.2022.108636
5. Chen J, Mei J, Li X, et al. TransUnet: rethinking the U-net archi- tecture design for medical image segmentation through the lens of transformers[J]. Med Image Anal. 2024;97:103280. doi:10.1016/j.media.2024.103280
6. Hedayati M, Kazerouni A, Soltany M, et al. Hiformer: hierarchical multi-scale representations using transformers for medical image segmentation[C]. Proceedings of the IEEE/CVF winter confer- ence on applications of computer vision. 2023: 6202-6212.