Learning with Explicit Topological Priors for Chest X-ray Rib Segmentation

Xiaowei Zhao[1,3], Chenglong Li[1,3]✉, Jin Tang[2,3], and Chuanfu Li[4]

[1]School of Artificial Intelligence, Anhui University, Hefei, China
[2]School of Computer Science and Technology, Anhui University, Hefei, China
[3]State Key Laboratory of Opto-Electronic Information Acquisition and Protection Technology
[4]The First Affiliated Hospital, Anhui University of Chinese Medicine, Hefei, China
lcl1314@foxmail.com

Abstract. Chest X-ray (CXR) image examination is a primary tool for assessing thoracic abnormalities. It is widely utilized for initial diagnosis and screening of diseases due to its cost-effectiveness and low radiation dose. Segmentation of ribs in CXR images (CXR rib segmentation) facilitates rapid determination of lesion types and locations, thereby alleviating the workload of medical professionals. Deep learning-based methods have achieved significant progress but still face some challenges in CXR rib segmentation, such as the occlusion challenge caused by artifacts and the interlace challenge caused by the spatial overlap of ribs. Therefore, it can be observed that the topological knowledge of ribs is crucial for CXR rib segmentation but neglected in existing methods, including the connectivity and interactivity of ribs. To address these challenges, we propose a novel learning framework that integrates explicit topological priors into segmentation networks for precise CXR rib segmentation. In particular, we introduce two modules including the connectivity prior embedding module and the interactivity prior embedding module. These modules are designed to explicitly encode the continuity and interactivity of ribs into deep learning models for end-to-end training. Both modules are plug-and-play and can be integrated into various networks. We conduct extensive experiments on VinDr-RibCXR and CXRS datasets to evaluate the segmentation accuracy of each rib using multiple metrics. Evaluation and visual results show that our method exhibits strong adaptability, seamlessly integrating with diverse architectures and enhancing performance across various networks. Our code is publicly available at https://github.com/XWei98/LTSeg.

Keywords: Chest X-ray Image · Rib Segmentation · Topological Prior.

1 Introduction

Chest X-ray (CXR) imaging, with its advantages of simplicity, fast imaging, low cost, and low radiation exposure [9], remains a primary modality for screening and diagnosing lung diseases such as COVID-19 [7]. Segmentation of ribs

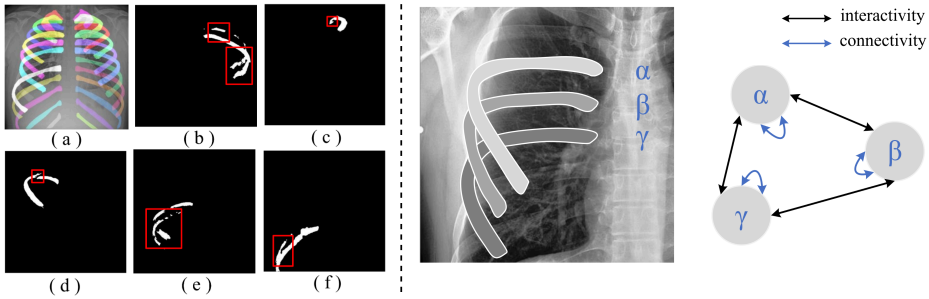

Fig. 1: Left: Incorrect segmentation results for different ribs. Right: Schematic representation of topological priors: connectivity and interactivity. Connectivity: α, β, γ, should all be complete individuals. Interactivity: α is above β and partially overlaps, γ is below β and partially overlaps.

is also essential in medical image analysis, which helps to improve lesion detection and disease diagnosis by quantitatively analyzing accurate segmentation results accuracy [4], personalized patient care [17]. Moreover, the stable structure and morphology of ribs serve as reliable references for various analytical and quantification tasks, including estimating lung volume [20,2], quantifying bone abnormalities, and assessing pediatric spinal deformities. However, as shown in Fig.1-left, CXR rib segmentation is a challenging task. (1) Due to the complex interleaving of the 24 ribs in the CXR image, artefacts can also arise from the contact of different structures with the surrounding tissue [8,13]. This may mislead the network into the surrounding ribs area, as shown in (b), (e), and (f). (2) Due to lesions in the lungs or other conditions causing abnormal local grayscale values in CXR image, the segmentation results of ribs may appear disjointed, as shown in (c), and (d). These challenges can adversely affect the accuracy of rib segmentation.

To solve these problems, inspired by previous research [3] we propose a new method to learn topological prior knowledge of interactions and connectivity, as shown in Fig.1-right. It can also be encoded into the training process of deep neural networks to help them learn autonomously. Specifically, we propose two topological constraint modules: the Connectivity prior Embedding Module (CEM) and the Interactivity prior Embedding Module (IEM). We use these two modules to encode the topological prior knowledge of partial overlap and connectivity present in the rib structure into a deep neural network, identify the pixel points that violate the topological prior knowledge, and apply violation penalties to these pixel points to make the network pay more attention to these locations.

Our main contributions are summarized as follows: 1)We propose an interactivity prior embedding module to constrain the interactive overlap between ribs and a connectivity prior embedding module to constrain the internal connectivity of ribs. 2)Both modules can be plug-and-played into any deep neural network to encode conforming topological prior knowledge into deep neural networks for

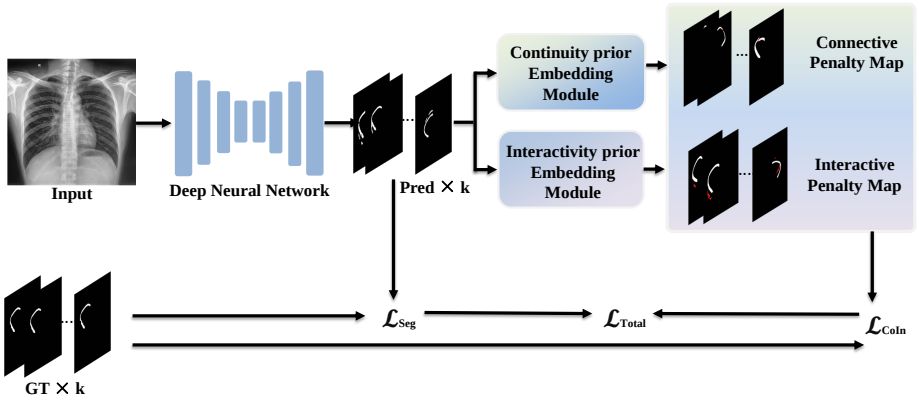

Fig. 2: The overall structure of our proposed framework. K represents the number of anatomical structures.

end-to-end training. 3)We conduct extensive experiments on multiple CXR rib segmentation datasets, CXRS and VinDr-RibCXR datasets. Both the evaluation and visualization results demonstrate the effectiveness of our method.

2 Methodology

From a morphological perspective, ribs in CXR image have two fixed topological priors: connectivity and interactivity. In Fig. 1-right, we use three rib labels α, β, and γ as specific examples to illustrate these topological priors. Connectivity implies that the ribs α, β, and γ should be complete entities. Interactivity refers to the crossing and partial overlap of α, β, and γ in the imaging of CXR image. Due to occlusion and overlap between different structures and large shadow regions, issues like missegmentation and incomplete segmentation arise during the segmentation process of individual ribs. For other ribs, we consider the current rib structure as β, the structure above it with interaction as α, and the structure below it with interaction as γ, and we observe that their topological relationships in the same way. Therefore, encoding topological priors such as connectivity and interactivity priors into deep neural networks for constraint is necessary.

2.1 Overview

The CXR image is input into a deep neural network to predict each of the 24 ribs individually. Next, each rib prediction and its ground-truth are processed by the Connectivity prior Embedding Module to identify violations of the continuity prior. Simultaneously, each rib, its upper and lower neighbors, and their respective ground-truths are input into the Interactivity Prior Embedding Module to detect violations of the interactivity prior. Regions with topological inconsistencies are penalized to enhance feature learning, as shown in Fig. 2. The proposed

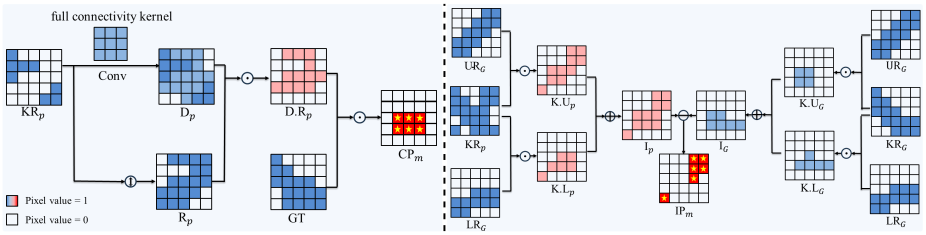

Fig. 3: The detail of our Connectivity prior Embedding Module (left) and Interactive prior Embedding Module (right).

modules are integrated into an end-to-end training framework. The network jointly optimizes the segmentation loss L_{Seg} and topological penalty L_{CoIn}, ensuring adherence to intrinsic topological priors while improving segmentation accuracy.

2.2 Connectivity Prior Embedding Module

The Connectivity prior Embedding Module (CEM) encodes the topological prior knowledge of the connectivity inside each rib in the CXR image into a deep neural network. Specifically, we first obtain binary masks for these predicted images separately and input the binary masks of each rib, denoted as the Key Rib Pred (KR_p), into the CEM. After that, we employ a manually designed $K \times K$ full connectivity convolutional kernel to fill the mask, thus filling the gaps in KR_p, resulting in the filled Dilation Pred (D_p). Fully connected convolution refers to setting all pixels within the current convolution kernel to foreground pixels as long as a foreground pixel appears. However, this convolution operation will cause unnecessary filling pixels to appear on the outside. In order to avoid the extra pixel filling, we first invert the foreground pixels and background pixels (denoted as $\updownarrow$) in the original Key Rib Pred to get the Reversal Pred (R_p). Subsequently, we multiply D_p with R_p to obtain the intersection region, denoted as $D.R_p$. Finally, we perform pixel-wise multiplication between $D.R_p$ and the binary ground-truth (GT) images, effectively excluding the genuine areas that do not violate the connectivity prior, yielding the final essential connectivity penalty map (CP_m). We can quantize the above steps as Equation 1:

$$\begin{aligned} D_p &= KR_p \circledast K, \qquad R_P = KR_p \updownarrow, \\ D.R_p &= D_p \odot R_p, \qquad CP_M = D.R_p \odot GT, \end{aligned} \tag{1}$$

the symbol $\circledast$ denotes the standard convolution operation, and K represents the convolution kernel of our fully connected kernel. $\odot$ denotes the Hadamard product, and $\updownarrow$ denotes the reversal of front and background pixels. The detailed process is visualized in Fig. 3-left. It condenses the computed connectivity constraint maps for each target into binary masks. This enables us to integrate topological connectivity prior knowledge into any deep neural network, constraining

the existing feature information with the Connectivity prior Embedding Module. Enabling the model to overcome the problem of incomplete segmentation of ribs due to occlusion and shaped shadows, which in turn improves the accuracy of rib segmentation.

2.3 Interactivity Prior Embedding Module

The Interactivity prior Embedding Module (IEM) encodes topological prior knowledge of interactivity between ribs in CXR image into a deep neural network. Specifically, we first obtain binary masks for these predicted images. Then, we simultaneously input the binary mask of each rib, denoted as Key Rib Pred (KR_p), and the ground-truth of the ribs above and below it, denoted as Upper Rib Ground-truth (UR_G) and Lower Rib GT (LR_G) into the Interactive Constraint Module. We perform pixel-wise multiplication between KR_P and UR_G to obtain their intersection, denoted as $K.U_p$. Similarly, pixel-wise multiplication between KR_P and LR_G results in their intersection, denoted as $K.L_p$. Then, we add these two binary masks $K.U_P$ and $K.L_p$ to obtain the Inter Pred (I_p). I_p represents the interference from surrounding ribs, mistakenly segmented as the pixel region of the current rib in the prediction of a single rib. However, due to the 2D projection of different ribs in the imaging process of the CXR image, there are naturally overlapping parts of the ribs in the CXR image. Therefore, to avoid imposing penalties on these naturally existing overlapping regions, we operate on the ground-truth of the key rib and its upper and lower ribs in the same way as before, obtaining Inter GT (I_G). I_G represents the naturally occurring overlap regions of the current rib with the surrounding ribs in the CXR image. We eliminate the naturally overlapping regions in Inter to obtain the Interactive Penalty Map (IP_m), which represents the violation of the topological prior knowledge of interactivity between ribs in the predicted results of the current rib. We can quantize the above steps as Equation 2:

$$\begin{aligned} K.U_p = KR_p \odot UR_G, \quad K.L_P = KR_p \odot LR_G, \\ I_p = K.U_p \oplus K.L_p, \quad IP_m = I_p \ominus I_G, \end{aligned} \tag{2}$$

where $\oplus$ denotes the union operation and $\odot$ denotes the Hadamard product. $\ominus$ indicates exclusion. The detailed process is visualized in Fig. 3-right. This enables us to integrate topological prior knowledge into any deep neural network, constraining the existing feature information with the Interactivity prior Embedding Module. Enabling the model to distinguish adjacent ribs improves the accuracy of rib segmentation.

2.4 Incorporating into End-to-End Training

To integrate our modules into an end-to-end training framework, we propose a topological prior knowledge constraint loss. Penalties are applied to the pixels identified by IEM and CEM that violate topological prior constraints, thereby correcting these inconsistencies. Specifically, let $F \in \mathbb{R}^{B\times C\times H\times W}$ represent the

multi-likelihood map predicted by the network, where B is the batch size, C is the number of segmentation categories, and H, W denote the height and width of the map. Similarly, $G \in \{0,1\}^{B\times C\times H\times W}$ is the ground-truth segmentation map. The original segmentation loss, L_{Seg}, employs Dice loss, a widely used metric in segmentation tasks. Additionally, L_{CoIn} represents the topological prior loss, imposing penalties on regions identified by CP_m and IP_m that violate prior constraints, as shown in Equation 3:

$$\begin{aligned} L_{CoIn} &= L_{tpc}(F \odot (CP_m + IP_m), G \odot (CP_m + IP_m)), \\ L_{All} &= L_{Seg} + \lambda L_{CoIn}. \end{aligned} \tag{3}$$

Our topological priori constraint loss L_{CoIn} is controlled by the weights λ and summed with L_{Seg} to get the overall loss L_{All}.

3 Experiment and Results

3.1 Datasets and Evaluation Metrics

To validate the effectiveness of our approach, we perform training and evaluation on the VinDr-RibCXR [10] and CXRS [5] rib segmentation datasets.

CXRS Dataset [5] consists of 1,242 CXR images, which we split into training, validation, and testsets (lesion images). Each image is annotated with pixel-level labels for 24 ribs, including the left side (L1–L12) and the right side (R1–R12). We redivided the dataset to ensure that the test set included pathological conditions of different degrees, providing a more challenging scenario for anatomical segmentation.

VinDr-RibCXR Dataset [10] contains 245 CXR images, with 196 allocated for training and the remaining 49 for testing. Each image includes segmentation annotations for 20 ribs, covering L1–L10 on the left and R1–R10 on the right.

Evaluation Metrics We employ standard evaluation metrics, including positive measures (IoU, DSC, Specificity, and Sensitivity) as well as negative measures: Average Symmetric Surface Distance (ASSD) and Hausdorff Distance (HD).

Experiment Settings All models were implemented on a GeForce RTX 4090 with an AMD EPYC 7542 32-Core Processor @ 3.164 GHz and 251 GB memory. We set the image input size to 448×448 and trained the network using the widely-used Adam optimizer [6]. We trained all the networks for 300 epochs. And we chose Dice loss as our primary loss function.

3.2 Experiment Results

We assess our method on three publicly available medical image segmentation datasets, benchmarking it against approaches built on different architectures, including CNN-based methods [11,19,15], Transformer-based models [1,16], Mamba-based frameworks [12], and SAM-based method [18].

We uniformly apply our Connectivity prior Embedding Module (CEM) and Interactivity prior Embedding Module (IEM) to the aforementioned methods,

Table 1: Quantitative comparison on the VinDr-RibCXR and CXRS. † denotes the method after loading the pre-trained model.

Methods	VinDr-RibCXR [10]						CXRS [5]					
	mIOU↑	mDSC↑	mSen↑	mSpec↑	mHD↓	mASSD↓	mIOU↑	mDSC↑	mSen↑	mSpec↑	mHD↓	mASSD↓
UNEXT [14]	42.75	57.90	**56.39**	99.42	46.06	5.447	44.76	59.01	**59.67**	99.35	41.71	5.571
UNEXT$_{ours}$	**43.95**	**59.30**	53.83	**99.52**	**38.09**	**4.481**	**45.77**	**60.08**	56.76	**99.44**	**37.75**	**5.250**
Unet++ [19]	59.25	72.49	**73.85**	99.58	27.67	2.792	58.02	70.51	70.55	**99.54**	30.70	4.177
Unet++$_{ours}$	**60.38**	**73.39**	72.96	**99.62**	**27.42**	**2.712**	**58.23**	**70.76**	**73.09**	99.51	**30.09**	**4.168**
Unet [11]	59.82	73.14	**74.24**	99.59	29.29	2.650	58.68	71.11	**73.42**	99.52	30.18	4.038
Unet$_{ours}$	**60.80**	**73.93**	72.39	**99.65**	**28.25**	**2.564**	**58.92**	**71.34**	71.01	**99.56**	**30.41**	**4.004**
UCtransUnet [16]	59.68	72.88	**72.83**	99.61	28.60	2.747	62.58	74.48	**75.41**	99.60	27.9	3.313
UCtransUnet$_{ours}$	**60.43**	**73.61**	71.87	**99.65**	**28.41**	**2.629**	**62.73**	**74.61**	73.29	**99.64**	**28.18**	**3.306**
AttenUnet [15]	63.47	75.97	**76.32**	99.65	26.28	2.275	63.21	74.75	**76.42**	99.58	26.13	3.472
AttenUnet$_{ours}$	**63.92**	**76.32**	74.73	**99.68**	**26.36**	**2.286**	**63.62**	**75.06**	74.88	**99.62**	**26.16**	**3.443**
VM-Unet† [12]	63.75	76.70	**78.1**	99.65	**21.45**	1.373	66.67	78.30	**79.28**	99.64	23.30	2.238
VM-Unet†$_{ours}$	**65.16**	**77.96**	76.52	**99.71**	22.00	**1.231**	**67.68**	**79.07**	77.32	**99.70**	**22.72**	**2.149**
TransUnet [1]	63.77	76.36	**76.78**	99.66	25.14	1.898	65.83	77.26	**78.20**	99.64	23.40	2.635
TransUnet$_{ours}$	**65.86**	**77.99**	76.60	**99.71**	**23.33**	**1.790**	**66.36**	**77.51**	75.97	**99.68**	**22.89**	**2.644**
MedSAM† [18]	65.48	78.65	78.23	99.76	19.74	1.188	67.65	79.45	**78.59**	99.71	15.57	1.642
MedSAM†$_{ours}$	**66.28**	**79.23**	**78.66**	**99.76**	**18.67**	**1.083**	**68.40**	**79.91**	78.57	**99.73**	**15.32**	**1.623**
TransUnet† [1]	70.38	81.37	**81.76**	99.73	19.70	1.334	69.83	80.23	**81.04**	99.69	21.07	2.286
TransUnet†$_{ours}$	**71.09**	**82.07**	80.65	**99.77**	**19.33**	**1.221**	**70.66**	**81.01**	80.47	**99.72**	**19.18**	**2.137**

Table 2: **The Left**: Ablation studies of the CEM and IEM on the CXRS dataset. **The Right**: Ablation Study for convolutional kernels of CEM on the CXRS dataset.

CEM	IEM	mIOU	mDSC	mSen	mSpec	mHD	mASSD
-	-	69.83	80.23	81.04	99.69	21.07	2.286
✓	-	70.06	80.48	**81.20**	99.69	20.12	2.243
-	✓	70.49	80.91	80.07	99.72	19.42	2.148
✓	✓	**70.66**	**81.01**	80.47	**99.72**	**19.18**	**2.137**

Conv kernels	mIOU	mDSC	mSen	mSpec	mHD	mASSD
3×3	70.40	80.80	79.30	**99.74**	21.60	2.207
5×5	**70.66**	**81.01**	**80.47**	99.72	**19.18**	**2.137**
7×7	70.44	80.80	79.46	99.73	19.73	2.201
1×5	70.27	80.63	80.00	99.72	20.07	2.233
5×1	70.58	80.92	80.43	99.72	19.32	2.111

incorporate the topological prior constraints for training, namely Unet$_{Ours}$, Unet++$_{Ours}$, AttUnet$_{Ours}$, UCTransUnet$_{Ours}$, , TransUnet$^{\dagger}_{Ours}$, VMUNet$^{\dagger}_{Ours}$ and MedSAM$^{\dagger}_{Ours}$. For the experiments based on SAM, we all adopt the prompt-free strategy to avoid relying on expert prompts. Experiments on the VinDr-RibCXR and CXRS datasets demonstrate that our method outperforms existing approaches across multiple metrics, as shown in Table 1, particularly in mIOU and mDSC. For instance, compared to TransUnet, our approach achieves a 1.12% improvement in mIOU and a 1.43% improvement in mDSC on the VinDr-RibCXR dataset, while also achieving state-of-the-art performance on the CXRS dataset. Furthermore, our method exhibits superior robustness in terms of mHD and mASSD. These results demonstrate that our method effectively embeds interaction and connectivity priors specific to ribs into the different architectures. The visualization results are shown in Fig. 4.

3.3 Ablation Study

To substantiate the efficacy of our proposed method, we conduct comprehensive ablation studies on the CXRS dataset based on TransUnet†, meticulously exam-

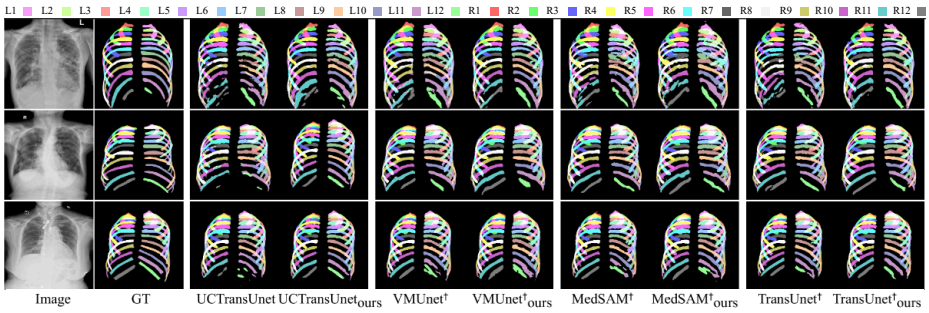

Fig. 4: Visualization results on the CXRS dataset.

Table 3: **The Left**: Ablation study on the topological prior constraints L_{tpc}. **The Right**: Ablation study on the weight λ.

L_{tpc}	mIOU	mDSC	mSen	mSpec	mHD	mASSD
None	69.83	80.23	**81.04**	99.69	21.07	2.286
Dice	**70.66**	**81.01**	80.47	99.72	**19.18**	**2.137**
MSE	70.30	80.65	80.98	99.70	19.12	2.193
BCE	70.40	80.80	79.30	**99.74**	21.60	2.207

λ	mIOU	mDSC	mSen	mSpec	mHD	mASSD
$\lambda = 0$	69.83	80.23	**81.04**	99.69	21.07	2.286
$\lambda = 0.1$	70.16	80.66	80.73	99.70	20.42	2.156
$\lambda = 0.2$	70.47	80.84	80.47	99.71	19.53	2.177
$\lambda = 0.3$	**70.66**	**81.01**	80.47	**99.72**	**19.18**	**2.137**
$\lambda = 0.4$	70.49	80.80	78.94	99.74	20.10	2.172

ining the impact of the Topology Prior Embedding Module CEM and IEM, the significance of different weights, and the influence of various loss functions.

Ablation Study for CEM and IEM. To further validate the effectiveness of our Connectivity prior Embedding Module (CEM) and Interactivity prior Embedding Module (IEM). As shown in Table 2-left, we can see that only embedding the CEM into the network for training has a positive impact on every rib structure. When we embed only the IEM, we observe a more substantial enhancement in performance compared to the baseline and CEM, with an increase of 0.66% in mIOU, an improvement of 0.68 in mDSC. Finally, we simultaneously embed both the IEM and CEM into the network, achieving the best results. These results clearly demonstrate the effectiveness of our modules.

Ablation Study for Kernels of CEM. To investigate the impact of the choice of fully connected convolutional kernels in the CEM on performance, we are using five types of kernels: 3×3, 5×5, 7×7, 1×5, and 5×1, as shown in Table 2-right. When extend the kernel to 5×5, we achieve the best result.

Ablation Study for Loss Functions. conduct ablation studies by employing various pixel-level loss functions for L_{tpc}, as shown in Table 3-left, where the "None" entry represents the outcome without the inclusion of the L_{tpc}. When employing Dice loss, we attain the optimal result.

Ablation Study for Loss Weights. We introduce a parameter λ, to control the balance of their learning, as shown in Equation 3. The experimental results are shown in Table 3-right, where the "0" entry represents the baseline without using L_{tpc}. When setting λ to 0.3, our method achieved optimal result.

4 Conclusion

In this study, we propose the CEM and IEM to encode fixed topological prior knowledge between ribs in CXR images as constraints into networks. Furthermore, we introduce a topological prior constraint loss, which penalizes segmentation regions that violate the topological prior knowledge. However, as CXR images are 2D projections of the chest's 3D structure, our method primarily focuses on 2D topological relationships, overlooking the spatial relationships lost during projection. Future work will explore more effective strategies to leverage the spatial information of ribs.

Acknowledgments. This work was supported by the Natural Science Foundation of China (No.62376004), the Innovation Program of Anhui Province (No.GXXT-2022-031) and the Major Natural Science Research Projects of the Anhui Provincial Department of Education (No.2023AH040102).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Chen, J., Mei, J., Li, X., Lu, Y., Yu, Q., Wei, Q., Luo, X., Xie, Y., Adeli, E., Wang, Y., et al.: Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers. Medical Image Analysis p. 103280 (2024). https://doi.org/10.1016/j.media.2024.103280
2. Gaggion, N., Mansilla, L., Mosquera, C., Milone, D.H., Ferrante, E.: Improving anatomical plausibility in medical image segmentation via hybrid graph neural networks: applications to chest x-ray analysis. IEEE Transactions on Medical Imaging **42**(2), 546–556 (2022)
3. Gupta, S., Hu, X., Kaan, J., Jin, M., Mpoy, M., Chung, K., Singh, G., Saltz, M., Kurc, T., Saltz, J., Tassiopoulos, A., Prasanna, P., Chen, C.: Learning topological interactions for multi-class medical image segmentation. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (eds.) Computer Vision – ECCV 2022. pp. 701–718. Springer Nature Switzerland, Cham (2022)
4. Hu, G., Kang, Y., Zhao, G., Jin, Z., Li, C., Tang, J.: Dynamic strip convolution and adaptive morphology perception plugin for medical anatomy segmentation. IEEE Transactions on Medical Imaging **44**(6), 2541–2552 (2025)
5. Huang, L., Ma, D., Zhao, X., Li, C., Zhao, H., Tang, J., Li, C.: Semantics guided disentangled gan for chest x-ray image rib segmentation. In: Pattern Recognition and Computer Vision. pp. 46–60. Springer Nature Singapore, Singapore (2025)
6. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
7. Li, F., Lu, X., Yuan, J.: Mha-corocapsule: Multi-head attention routing-based capsule network for covid-19 chest x-ray image classification. IEEE Transactions on Medical Imaging **41**(5), 1208–1218 (2022). https://doi.org/10.1109/TMI.2021.3134270
8. Loog, M., van Ginneken, B., Schilham, A.: Filter learning: Application to suppression of bony structures from chest radiographs. Medical Image Analysis **10**(6), 826–840 (2006)

9. Ma, D., Pang, J., Gotway, M.B., Liang, J.: A fully open ai foundation model applied to chest radiography. Nature (2025). https://doi.org/10.1038/s41586-025-09079-8
10. Nguyen, H.C., Le, T.T., Pham, H.H., Nguyen, H.Q.: Vindr-ribcxr: A benchmark dataset for automatic segmentation and labeling of individual ribs on chest x-rays. arXiv preprint arXiv:2107.01327 (2021)
11. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer International Publishing, Cham (2015)
12. Ruan, J., Li, J., Xiang, S.: Vm-unet: Vision mamba unet for medical image segmentation. arXiv preprint arXiv:2402.02491 (2024)
13. Suzuki, K., Abe, H., MacMahon, H., Doi, K.: Image-processing technique for suppressing ribs in chest radiographs by means of massive training artificial neural network (mtann). IEEE Transactions on Medical Imaging **25**(4), 406–416 (2006)
14. Valanarasu, J.M.J., Patel, V.M.: Unext: Mlp-based rapid medical image segmentation network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 23–33. Springer (2022)
15. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, , Polosukhin, I.: Attention is all you need. Advances in Neural Information Processing Systems **30** (2017)
16. Wang, H., Cao, P., Wang, J., Zaiane, O.R.: Uctransnet: Rethinking the skip connections in u-net from a channel-wise perspective with transformer. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 2441–2449 (2022)
17. Wang, X., Li, Y., Wu, W., Jin, J., Rong, Y., Jiang, B., Li, C., Tang, J.: Pre-training on high-resolution x-ray images: An experimental study. Visual Intelligence **3**(1), 8 (2025)
18. Wu, J., Wang, Z., Hong, M., Ji, W., Fu, H., Xu, Y., Xu, M., Jin, Y.: Medical sam adapter: Adapting segment anything model for medical image segmentation. Medical Image Analysis **102**, 103547 (2025)
19. Zhou, Z., Siddiquee, M., Tajbakhsh, N., Liang, J.: Redesigning skip connections to exploit multiscale features in image segmentation. arXiv preprint arXiv:1912.05074 (2020)
20. Çallı, E., Sogancioglu, E., van Ginneken, B., van Leeuwen, K.G., Murphy, K.: Deep learning for chest x-ray analysis: A survey. Medical Image Analysis **72**, 102125 (2021)