

Topology-Constrained Learning for Efficient Laparoscopic Liver Landmark Detection

Ruize Cui¹, Jiaan Zhang², Jialun Pei^{3(✉)}, Kai Wang⁴, Pheng-Ann Heng³, and Jing Qin¹

¹ The Hong Kong Polytechnic University, Hong Kong, China

² National University of Singapore, Singapore

³ The Chinese University of Hong Kong, Hong Kong, China

⁴ Nanfang Hospital, Southern Medical University, Guangzhou, China
jialunpei@cuhk.edu.hk

Abstract. Liver landmarks provide crucial anatomical guidance to the surgeon during laparoscopic liver surgery to minimize surgical risk. However, the tubular structural properties of landmarks and dynamic intraoperative deformations pose significant challenges for automatic landmark detection. In this study, we introduce TopoNet, a novel topology-constrained learning framework for laparoscopic liver landmark detection. Our framework adopts a snake-CNN dual-path encoder to simultaneously capture detailed RGB texture information and depth-informed topological structures. Meanwhile, we propose a boundary-aware topology fusion (BTF) module, which adaptively merges RGB-D features to enhance edge perception while preserving global topology. Additionally, a topological constraint loss function is embedded, which contains a center-line constraint loss and a topological persistence loss to ensure homotopy equivalence between predictions and labels. Extensive experiments on L3D and P2ILF datasets demonstrate that TopoNet achieves outstanding accuracy and computational complexity, highlighting the potential for clinical applications in laparoscopic liver surgery. Our code is available at <https://github.com/cuiruize/TopoNet>.

Keywords: Landmark detection · Laparoscopic liver surgery · Topology constraint · RGB-D fusion.

1 Introduction

Laparoscopic liver surgery has been widely adopted due to its perioperative advantages of reduced blood loss, faster recovery, and lower complication rates [8,9]. However, the limited laparoscopic field of view and intraoperative deformations make the identification of key liver anatomical structures particularly challenging. Augmented reality (AR) navigation technology has emerged as a promising solution to provide visual guidance to surgeons by establishing correspondences between intraoperative 2D keyframes and preoperative 3D anatomy [1,10]. Liver landmarks, including silhouette, falciform ligament, and ridge, have been validated as informative biomarkers to provide consistent clues in preoperative-to-intraoperative relations and assist in decision-making [14–16]. In this regard,

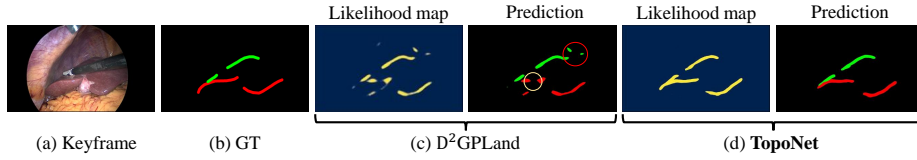


Fig. 1: Illustration of the crucial role of topological priors in tubular characteristics learning and topology preservation. We highlight a broken landmark in yellow circle and two topological false positives caused by outliers in red circle.

accurate liver landmark detection is essential for performing high-quality registration with preoperative 3D models and reaching AR navigation. However, automatic intraoperative liver landmark detection remains challenging due to liver deformation, varying laparoscopic viewpoints, and occlusions from outliers like surgical instruments. As a result, there is still demand for the development of intelligent computer-assisted techniques for robust liver landmark detection in a complex laparoscopic environment.

Compared to landmark definitions like points [25], bounding boxes [21], or contours [6], defining liver marker points as continuous semantic regions has been proven to be of great value in enhancing intraoperative spatial relationships [1, 16, 19]. Several existing studies have explored using deep-learning-based models to automatically segment landmark regions. Labrunie *et al.* [15] employed an original U-Net [18] to detect 2D landmarks and then aligned them with preoperative 3D models. Subsequently, more advanced networks [1], *e.g.*, nnU-Net [13] and UNet++ [28], have been applied for more accurate landmark detection. However, these methods primarily utilize existing segmentation frameworks while overlooking the anatomical characteristics of liver landmarks. More recently, D²GPLand [16] integrates depth geometric priors and prompt-guided training to enhance the learning of liver landmark features, further improving the detection results. Despite improvements in detection performance, the fine-tubular structural properties of liver landmarks and the presence of outliers still seriously affect the accuracy of landmark detection, as shown in Fig. 1 (c). Most existing approaches concentrate on pixel-level textural and geometric clues, ignoring the intrinsic topology characteristics of liver landmarks, which are effective in mitigating false positive detections from outlier occlusions and preserving the structural continuity of tubular landmarks. To this end, we consider embedding topological constraints into the network to develop an efficient and precise laparoscopic liver landmark detection algorithm.

In this work, we introduce a topology-constrained learning network, termed *TopoNet*, for efficient laparoscopic liver landmark detection. Specifically, considering that depth modality has been demonstrated to provide effective auxiliary geometric information [16], we design a **snake-CNN dual-path encoder**, which employs the proposed snake topology acquisition (STA) blocks and a CNN to extract depth topological structures and detail texture features, respectively. Then, a **boundary-aware topological fusion (BTF)** module is proposed for

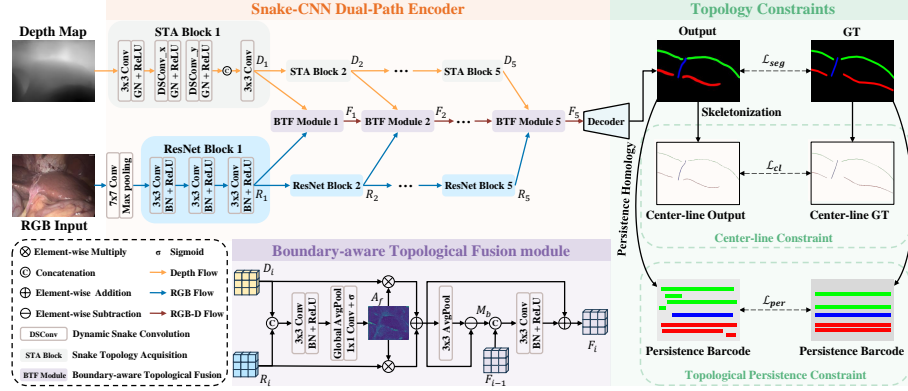


Fig. 2: Overall architecture of the proposed TopoNet.

adaptively merging RGB-D features and sensing boundary features, which facilitates the preservation of the global topology of liver landmarks (see likelihood maps in Fig. 1). To adequately exploit topology characteristics for detection supervision, we also present a **topological constraint loss function** composed of a center-line constraint loss and a topological persistence loss to ensure homotopy equivalence between predictions and labels. Experimental results on L3D and P2ILF datasets [1, 16] show that TopoNet outperforms 12 advanced models on the liver landmark detection task.

2 Methodology

The detailed pipeline of TopoNet is illustrated in Fig. 2. Our framework takes RGB keyframes and the corresponding depth maps estimated by the frozen AdelaiDepth [26] as inputs. We embrace multi-view snake convolutions into the snake topology acquisition (STA) block to capture deep topological cues and combine ResNet [11] to extract multiscale RGB features. Then, the depth feature D_i and RGB feature R_i output from corresponding encoder blocks are passed into the boundary-aware topological fusion (BTF) module to generate the fused features F_i . Notably, the first BTF module takes D_1 and R_1 as inputs, and the other ones utilize only the former fused feature F_{i-1} for residual learning to reduce information loss. Finally, we adopt a CNN decoder to produce landmark results and supervise our network with the proposed topological constraint loss to constrain the connectivity and topological persistence of detection maps.

2.1 Snake-CNN Dual-Path Encoder

To perceive semantic landmarks with slender and tortuous tubular structures, we design a snake-CNN dual-path encoder that separately extracts depth topology information and RGB texture clues with our STA blocks and CNN blocks.

As illustrated in Fig. 2, we employ a ResNet-34 encoder [11] as the backbone for RGB keyframes, *i.e.*, the i -th ResNet block outputs the RGB feature R_i . The depth pathway consists of five cascaded STA blocks for perceiving topological properties. In i -th STA block, the input depth feature D_{i-1} is first passed through a convolution block for lower-level feature acquisition. Then, we apply multi-view dynamic snake convolution (DSConv) [17] blocks to extract critical topology cues in the X- and Y-axes. Lastly, we concatenate single-axis features and use another convolution block for cross-axis feature aggregation to generate the output depth feature D_i .

2.2 Boundary-aware Topological Fusion

The proposed BTF module aims to integrate RGB and depth features effectively to capture comprehensive landmark features and preserve topological structures. As shown in Fig. 2, the input dual-modal features R_i and D_i are first concatenated along the channel dimensions and fed into the convolution block to obtain merged features. Then, a global average pooling operation and Sigmoid activation are applied to generate the fused attention map A_f that highlights the important regions. The fusion process can be formulated as

$$A_f = \sigma(\text{Pg}(\text{ReLU}(\text{BN}(\text{Conv}(\mathcal{C}[D_i, R_i]))))), \quad (1)$$

where Conv represents the convolution operation, BN is batch normalization. Pg is the global average pooling operation. Afterward, we multiply A_f by R_i, D_i and element-wise sum them to get the primary fused feature \hat{F}_i :

$$\hat{F}_i = (R_i \cdot A_f) \oplus (D_i \cdot A_f), \quad (2)$$

where \oplus stands for the element-wise summation. Upon getting \hat{F}_i , we design a boundary enhancement operation to guide the model to focus on the ambiguous edge regions. Here the principle is that liver landmarks are ultimately the boundary areas of the liver, and focusing on these regions is beneficial for holistic learning of anatomical structures, thus preserving the global topology of liver landmarks, *i.e.*, the relationships among different types of landmarks. Concretely, we apply an average pooling with 3×3 kernel and compute the feature differences by element-wise subtraction to obtain the boundary map M_b . After that, we perform multi-scale aggregation of M_b and the output of the previous BTF module F_{i-1} with convolution operation and add \hat{F}_i in the form of residual to reduce the information loss, resulting in the output F_i :

$$F_i = \hat{F}_i \oplus \text{Conv}(\text{BN}(\text{ReLU}(\mathcal{C}[(\hat{F}_i - \text{Po}(\hat{F}_i)), F_{i-1}]))), \quad (3)$$

where Po indicates the average pooling operation.

2.3 Topological Constraint Loss

To further refine the topological structure, we introduce the topological constraint loss function for supervision, involving a center-line constraint loss and a topological persistence loss.

Center-line Constraint Loss. Since tubular structures rely heavily on connectivity, traditional pixel-wise segmentation losses (*e.g.*, Dice Loss) that mainly focus on overlapping areas usually overlook structural discontinuities. To this end, we inject the center-line supervision into the loss function for ensuring landmark continuity. This supervision extends the cLDice loss function [20] to a multi-class form to supervise our network. Given the prediction P_l of landmark class l and its label G_l , we can extract their skeletons S_p^l and S_g^l . The topology precision T_{prec} and topology sensitivity T_{sens} can be defined as

$$T_{prec}(S_p^l, G_l) = \frac{|S_p^l \cap G_l|}{|S_p^l|}, \quad T_{sens}(S_g^l, P_l) = \frac{|S_g^l \cap P_l|}{|S_g^l|}. \quad (4)$$

Then we can compute the multi-class center-line constrained loss \mathcal{L}_{cl} :

$$\mathcal{L}_{cl} = \frac{1}{L} \sum_{l=1}^L \left(2 \times \frac{T_{prec}(S_p^l, G_l) \times T_{sens}(S_g^l, P_l)}{T_{prec}(S_p^l, G_l) + T_{sens}(S_g^l, P_l)} \right), \quad (5)$$

where $L = 3$ denotes the number of landmark classes to be detected.

Topological Persistence Loss. Impacted by outliers present in laparoscopic scenes, including instruments, blood, and similarly textured tissues, the network is prone to produce topological false positive results. To maintain the topological consistency between predictions and labels, we introduce a novel topological persistence loss \mathcal{L}_{per} based on the persistent homology theory [7]. Given a predictive likelihood map Y_l of landmark class l and its ground truth G_l , we exploit the efficient barcode computation algorithm [22] to obtain the persistence barcodes of Y_l and G_l in order to obtain the sets of birth and death coordinates of the matched connected components (denoted by B^m and D^m) and the unmatched ones (denoted by B^u and D^u). For each matched connected component C , we can find the persistence interval $C_{pre} = (B_{pc}^m, D_{pc}^m)$ in prediction from B^m and D^m , while the persistence interval in GT is denoted by $C_{gt} = (B_{gc}^m, D_{gc}^m)$. Here, we can obtain the persistence loss for the matched connected components by calculating the difference of the persistence intervals of C in prediction and GT:

$$\mathcal{L}_m^l = \frac{1}{N_m^l + s} \cdot \sum_{i \in M} (\|B_{pc}^m - B_{gc}^m\|_2 + \|D_{pc}^m - D_{gc}^m\|_2), \quad (6)$$

where N_m^l is the number of matched connected components of class l and $s = 1e^{-5}$ denotes the smoothing factor. M is the set of matched components. For each unmatched connected component Z in prediction, we can also get its persistence interval by $Z_{pre} = (B_{pz}^u, D_{pz}^u)$. Then we can compute the persistence loss for the unmatched components by calculating the length of its persistence interval:

$$\mathcal{L}_u^l = \frac{1}{N_u^l + s} \cdot \sum_{i \in U} \|B_{pz}^u - D_{pz}^u\|_2, \quad (7)$$

where N_u^l denotes the number of unmatched connected components of l and U is the set of unmatched components in prediction. The final L_{per} is derived by summing the individual losses and expanding to all landmark classes:

$$\mathcal{L}_{per} = \frac{1}{L} \sum_{l=1}^L \left(\frac{N_m^l}{N_m^l + N_u^l} \cdot \mathcal{L}_m^l + \frac{N_u^l}{N_m^l + N_u^l} \cdot \mathcal{L}_u^l \right). \quad (8)$$

Lastly, we combine two topological constraint losses with Dice loss \mathcal{L}_{dice} to reach the total loss function:

$$\mathcal{L}_{total} = \lambda_d \cdot \mathcal{L}_{dice} + \lambda_{cl} \cdot \mathcal{L}_{cl} + \lambda_{per} \cdot \mathcal{L}_{per}, \quad (9)$$

where $\lambda_d, \lambda_{cl}, \lambda_{per}$ are the balancing parameters for each loss function.

3 Experiments

3.1 Datasets and Metrics

We conduct experiments on two laparoscopic liver landmark datasets: L3D [16] and P2ILF [1]. L3D contains three liver landmark categories and consists of 1,152 annotated keyframes of 1920×1080 pixels from liver surgical videos. It is divided into three sets, where 921 frames from 32 patients for training, 122 frames from 3 patients for validation, and 109 frames from 4 patients for testing. P2ILF has 183 annotated laparoscopic images for liver landmark detection. Wherein, 167 images from 10 patients are used for training while the others are used for testing. Since only the training set is publicly available, we randomly select 124 images from 8 patients for training and the remaining for testing.

For evaluation metrics, we follow the experimental setup in [16], utilizing the Intersection over Union (IoU), Dice Score Coefficient (DSC), and Average Symmetric Surface Distance (Assd). We also compute the average inference speed and GFLOPs to evaluate the model efficiency.

3.2 Implementation Details

The training and testing processes of TopoNet are executed on a single NVIDIA RTX A6000 GPU. We train 100 epochs with a batch size of 4. We employ the Adam optimizer with the initial learning rate of $8e-5$ and weight decay factor of $3e-5$. In addition, the CosineAnnealingLR scheduler is used to adjust the learning rate to $1e-6$ at the end of the training. A warmup strategy is also applied to adaptively embed topology constraints. We empirically set $\lambda_d = 0.4$, $\lambda_{cl} = 0.4$, and $\lambda_{per} = 0.2$ for balancing parameters.

3.3 Comparison with State-of-the-art Methods

We follow the experimental settings of L3D benchmark to compare TopoNet with 12 cutting-edge methods. As shown in Table 1, TopoNet achieves more precise

Table 1: Comparison with cutting-edge methods on L3D and P2ILF test sets. We use **bold** and underline to mark the best and the second best results, respectively.

Methods	L3D [16]			P2ILF [1]			Infer. Speed ↓	GFLOPs ↓
	DSC ↑	IoU ↑	Assd ↓	DSC ↑	IoU ↑	Assd ↓		
U-Net [18]	51.39	36.35	84.94	29.89	17.89	47.95	172.65ms	774.30
COSNet [14]	56.24	40.98	69.22	26.78	16.05	47.33	167.25ms	319.24
Res-UNet [24]	55.47	40.68	70.66	25.84	15.76	50.97	155.76ms	314.65
UNet++ [28]	57.09	41.92	74.31	34.96	21.57	42.39	213.19ms	554.87
HRNet [23]	58.36	43.50	70.02	33.64	21.31	44.16	146.82ms	376.52
TransUNet [4]	56.81	41.44	76.16	25.49	15.22	52.14	262.21ms	795.10
Swin-UNet [2]	57.35	42.09	72.80	18.65	10.62	68.46	<u>127.22ms</u>	129.68
SAM-Adapter [5]	57.57	42.88	74.31	21.12	12.00	57.13	300.21ms	489.35
SAMed [27]	62.03	47.17	61.55	31.73	19.42	40.08	278.56ms	488.95
SAM-LST [3]	60.51	45.03	68.87	28.75	18.38	46.53	304.78ms	517.68
AutoSAM [12]	59.12	44.21	62.49	25.71	15.43	53.65	337.44ms	589.21
D ² GPLand [16]	<u>63.52</u>	<u>48.68</u>	<u>59.38</u>	<u>40.55</u>	<u>25.87</u>	<u>38.73</u>	297.93ms	572.85
TopoNet (Ours)	65.19	50.56	28.07	41.36	26.88	30.16	86.43ms	<u>276.99</u>

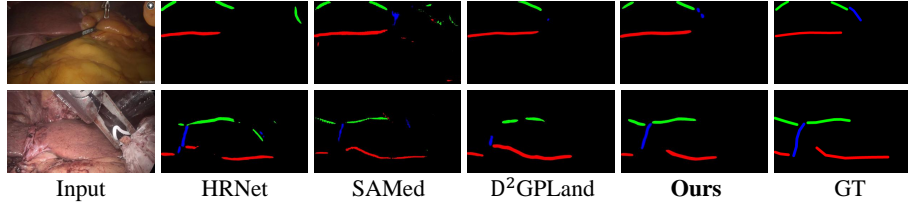


Fig. 3: Visual comparison of our TopoNet with representative models.

detection results on both datasets than other models. In addition, our results are statistically significant based on significance tests ($p\text{-value} < 0.05$). Compared to the second-ranked D²GPLand, TopoNet improves 1.67% on DSC, 1.88% on IoU, and **31.31** pixels on Assd on L3D dataset. For the P2ILF dataset, our method improves the performance on the DSC, IoU, and Assd metrics by 0.81%, 1.01%, and **8.57** pixels, respectively. Notably, we observe that the performance improvement in the Assd metric is tremendous. Assd is defined to compute the shortest distance from each foreground pixel in prediction to the foreground in GT. As mentioned in Sec. 2.3, the outliers in surgical keyframes affect the model to produce topological false positives in prediction, which are not typically located in locations that overlap with the target landmarks, and each pixel in these topologically inaccurate predictions contributes significantly to the Assd metric. The topological constraint loss, especially L_{per} , proposed in our method can suppress the negative effects of these outliers and preserve the topology structure, resulting in a substantial improvement in the Assd metric.

For model efficiency, the inference speed of TopoNet outperform all compared models and the computational complexity of our model also reaches the leading level. Compared to D²GPLand, the inference speed of TopoNet is about four

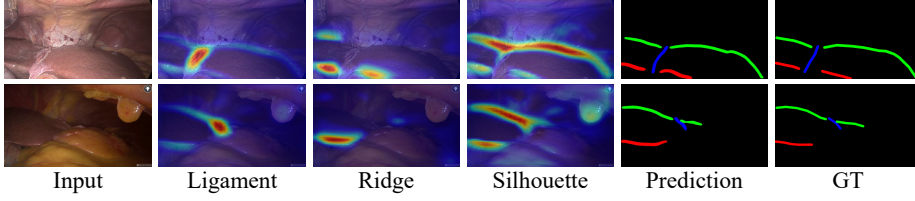


Fig. 4: Visualizations of class-aware attention maps with our predictions.

Table 2: Ablations for key Designs.

Methods	DSC	IoU	Assd
Baseline	54.64	40.94	49.64
w/o \mathcal{L}_{per}	58.87	46.69	38.66
w/o \mathcal{L}_{cl}	58.63	46.37	32.01
w/o \mathcal{L}_{per} & \mathcal{L}_{cl}	57.44	45.86	43.32
w/o BTF	57.92	46.03	37.47
TopoNet	59.79	47.38	29.27

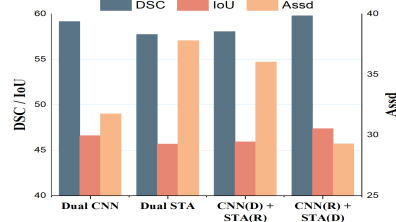


Fig. 5: Ablations for backbones.

times faster while requiring only about half the GFLOPs. Fig. 3 also exhibits the visualizations of TopoNet and competitors. With topological characteristics, our method produces superior results while alleviating the influences of outliers.

3.4 Ablation Analysis

We conduct ablation studies on the key components of our model, including the boundary-aware topological fusion (BTF) module, center-line constraint loss \mathcal{L}_{cl} , and topological persistence loss \mathcal{L}_{per} . We use the evaluation set of L3D for experiments. In our baseline, we replace the BTF module with simple RGB-D concatenation and remove the proposed topological losses. As illustrated in Table 2, the topological losses contribute to the detection performance, and the addition of \mathcal{L}_{per} brings a significant improvement in Assd. When embedding the BTF module for comprehensive RGB-D fusion, there is also an improvement in detection performance with 1.87% in DSC, 1.35% in IoU, and 8.20 pixels in Assd. In short, both our BTF module and topological constraint loss functions play an indispensable role in optimizing the model performance.

We also analyze the effect of different backbones on the snake-CNN encoder in RGB-D feature extraction. As shown in Fig. 5, when applying ResNet-34 blocks for RGB keyframes and the STA blocks for depth maps, our framework achieves optimal performance in all metrics. We also display the attention maps from the BTF Module 5 (refer to Fig. 2) for two typical samples in Fig. 4. It can be seen that our TopoNet enables accurate detection of landmarks and precise identification of landmark-related regions from a topological standpoint.

4 Conclusion

This study presents a lightweight topological-constrained framework TopoNet for precise and efficient laparoscopic liver landmark detection. Our framework comprises a snake-CNN dual-path encoder for RGB-D feature extraction, coupled with a boundary-aware fusion module to integrate bi-modal features and preserve the topology. More importantly, a topological constraint loss is introduced to enhance the learning of topological characteristics of liver landmarks and prevent topological errors. Our method outperforms current state-of-the-art methods with faster inference. This work opens up new possibilities for precise and efficient liver landmark detection and facilitates the application in clinical surgery.

Acknowledgments. The work described in this paper was supported in part by a Shenzhen-Hong Kong-Macao Science and Technology Plan Project (Category C Project) under Shenzhen Municipal Science and Technology Innovation Commission (project no. SGDX20230821092359002), in part by the Research Grants Council of the Hong Kong Special Administrative Region, China, under Project T45-401/22-N and AoE/E-407/24-N, and in part by the Regional Joint Fund of Guangdong (Guangdong-Hong Kong-Macao Research Team Project) under Grant 2021B1515130003.

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Ali, S., Espinel, Y., Jin, Y., Liu, P., Güttner, B., Zhang, X., Zhang, L., Dowrick, T., Clarkson, M.J., Xiao, S., et al.: An objective comparison of methods for augmented reality in laparoscopic liver resection by preoperative-to-intraoperative image fusion from the MICCAI2022 challenge. *MedIA* **99**, 103371 (2025)
2. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: *ECCV* (2022)
3. Chai, S., Jain, R.K., Teng, S., Liu, J., Li, Y., Tateyama, T., Chen, Y.w.: Ladder fine-tuning approach for sam integrating complementary network. *Procedia Computer Science* **246**, 4951–4958 (2024)
4. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306* (2021)
5. Chen, T., Zhu, L., Deng, C., Cao, R., Wang, Y., Zhang, S., Li, Z., Sun, L., Zang, Y., Mao, P.: Sam-adapter: Adapting segment anything in underperformed scenes. In: *IEEE ICCV*. pp. 3367–3375 (2023)
6. Collins, T., Pizarro, D., Gasparini, S., Bourdel, N., Chauvet, P., Canis, M., Calvet, L., Bartoli, A.: Augmented reality guided laparoscopic surgery of the uterus. *IEEE TMI* **40**(1), 371–380 (2020)
7. Edelsbrunner, H., Harer, J.: *Computational topology: an introduction*. American Mathematical Soc. (2010)
8. Fretland, Å.A., Dagenborg, V.J., Bjørnelv, G.M.W., Kazaryan, A.M., Kristiansen, R., Fagerland, M.W., Hausken, J., Tønnessen, T.I., Abildgaard, A., Barkhatov, L., et al.: Laparoscopic versus open resection for colorectal liver metastases: the oslo-comet randomized controlled trial. *Annals of surgery* **267**(2), 199–207 (2018)

9. Giannone, F., Cassese, G., Del Basso, C., Alagia, M., Palucci, M., Sangiuolo, F., Panaro, F.: Robotic versus laparoscopic liver resection for difficult posterosuperior segments: a systematic review with a meta-analysis of propensity-score matched studies. *Surgical Endoscopy* pp. 1–13 (2024)
10. Haouchine, N., Dequidt, J., Peterlik, I., Kerrien, E., Berger, M.O., Cotin, S.: Image-guided simulation of heterogeneous tissue deformation for augmented reality during hepatic surgery. In: 2013 IEEE ISMAR. pp. 199–208 (2013)
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE CVPR. pp. 770–778 (2016)
12. Hu, X., Xu, X., Shi, Y.: How to efficiently adapt large segmentation model (sam) to medical images. *arXiv preprint arXiv:2306.13731* (2023)
13. Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P.F., Kohl, S., Wasserthal, J., Koehler, G., Norajitra, T., Wirkert, S., et al.: nnu-net: Self-adapting framework for u-net-based medical image segmentation. *arXiv preprint arXiv:1809.10486* (2018)
14. Labrunie, M., Pizarro, D., Tilmant, C., Bartoli, A.: Automatic 3d/2d deformable registration in minimally invasive liver resection using a mesh recovery network. In: MIDL (2023)
15. Labrunie, M., Ribeiro, M., Mourthadhoi, F., Tilmant, C., Le Roy, B., Buc, E., Bartoli, A.: Automatic preoperative 3d model registration in laparoscopic liver resection. *IJCARS* **17**(8), 1429–1436 (2022)
16. Pei, J., Cui, R., Li, Y., Si, W., Qin, J., Heng, P.A.: Depth-driven geometric prompt learning for laparoscopic liver landmark detection. In: MICCAI. pp. 154–164 (2024)
17. Qi, Y., He, Y., Qi, X., Zhang, Y., Yang, G.: Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation. In: IEEE ICCV. pp. 6070–6079 (2023)
18. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234–241 (2015)
19. Schneider, C., Allam, M., Stoyanov, D., Hawkes, D., Gurusamy, K., Davidson, B.: Performance of image guided navigation in laparoscopic liver surgery—a systematic review. *Surgical Oncology* **38**, 101637 (2021)
20. Shit, S., Paetzold, J.C., Sekuboyina, A., Ezhov, I., Unger, A., Zhylka, A., Pluim, J.P., Bauer, U., Menze, B.H.: cldice-a novel topology-preserving loss function for tubular structure segmentation. In: IEEE CVPR. pp. 16560–16569 (2021)
21. Smithmaitrie, P., Khaonualsri, M., Sae-Lim, W., Wangkulangkul, P., Jearanai, S., Cheewatanakornkul, S.: Development of deep learning framework for anatomical landmark detection and guided dissection line during laparoscopic cholecystectomy. *Heliyon* **10**(3) (2024)
22. Stucki, N., Bürgin, V., Paetzold, J.C., Bauer, U.: Efficient betti matching enables topology-aware 3d segmentation via persistent homology. *arXiv preprint arXiv:2407.04683* (2024)
23. Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., et al.: Deep high-resolution representation learning for visual recognition. *IEEE TPAMI* **43**(10), 3349–3364 (2020)
24. Xiao, X., Lian, S., Luo, Z., Li, S.: Weighted res-unet for high-quality retina vessel segmentation. In: ITME. pp. 327–331 (2018)
25. Yao, Q., Quan, Q., Xiao, L., Kevin Zhou, S.: One-shot medical landmark detection. In: MICCAI. pp. 177–188 (2021)
26. Yin, W., Zhang, J., Wang, O., Niklaus, S., Chen, S., Liu, Y., Shen, C.: Towards accurate reconstruction of 3d scene shape from a single monocular image. *IEEE TPAMI* (2022)

27. Zhang, K., Liu, D.: Customized segment anything model for medical image segmentation. arXiv preprint arXiv:2304.13785 (2023)
28. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE TMI* **39**(6), 1856–1867 (2019)