Attenuation-Resilient Alternating Optimization for Laparoscopic Liver Landmark Detection

Lanqing Liu[1], Ruize Cui[1], Jialun Pei[2(✉)], Diandian Guo[2], Tiffany Y. So[2], Pheng-Ann Heng[2], and Jing Qin[1]

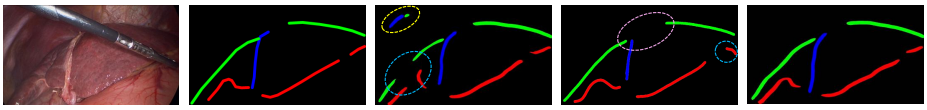
[1] The Hong Kong Polytechnic University, Hong Kong, China
[2] The Chinese University of Hong Kong, Hong Kong, China
peijialun@gmail.com

Abstract. Liver surface landmark detection is a fundamental prerequisite for anatomical guidance in laparoscopic liver surgery. However, it remains unreliable in practice due to two pervasive challenges: illumination attenuation in underexposed regions and the structural mismatch between pixel-wise localization and continuous curvilinear geometry. To address these limitations, we propose A2ONet, an attenuation-resilient alternating optimization network for robust liver landmark detection. To mitigate illumination attenuation, A2ONet embraces an illumination field compensation (IFC) block that adaptively enhances dark regions while preserving structural consistency. Meanwhile, we introduce a lightweight frequency–orientation selective filter (FOSF) to suppress repetitive texture interference and preserve salient curvilinear cues. Building upon these resilient representations, we design an alternating seg–curve optimization (ASCO) decoder that iteratively couples dense segmentation with explicit curve modeling, enabling mutual guidance to optimize both structural continuity and endpoint localization. Extensive evaluations on L3D-2K, L3D, and P2ILF demonstrate consistent improvements over competitive methods, establishing a more reliable foundation for intraoperative anatomy guidance. Our code is available at https://github.com/hyperiondk115/A2ONet.

Keywords: Landmark detection · Laparoscopic liver surgery · Parametric curve modeling · Low-light enhancement.

1 Introduction

Laparoscopic liver surgery offers substantial benefits in reduced trauma and faster postoperative recovery [18,8]. However, the restricted field of view and absence of tactile feedback significantly complicate intraoperative anatomical orientation. To mitigate these limitations, augmented reality (AR) guidance has been increasingly explored to link intraoperative laparoscopic views with preoperative CT/MRI-derived liver models [13,1,28]. Reliable detection of anatomical landmarks on the liver surface, *i.e.*, ridges, falciform ligaments, and silhouette boundaries, serves as semantic anchors for establishing 2D–3D correspondence



(a) Keyframe (b) Ground Truth (c) D^{2}GPLand+ (d) BCRNet (e) **A2ONet**

Fig. 1. Illustration of key challenges for liver landmark detection. Blue circles indicate errors caused by illumination attenuation, yellow circles denote segmentation false positives, and pink circles mark inaccurate endpoint localization.

and supporting accurate registration [20]. Consequently, anatomically consistent and geometrically reliable landmark detection is fundamental for robust AR-guided laparoscopic interventions.

Most existing studies formulate liver landmark detection as a dense segmentation problem [21,24,1]. In real surgical scenarios, however, there exists a structural mismatch between pixel-level targets and the continuous-curve geometry of anatomical landmarks. Although mask-based segmentation methods achieve remarkable performance in pixel-level localization, they often fragment delicate curvilinear landmarks into disconnected pieces, yielding local false positives. Recent studies [19,6] leverage large vision models [12] to enhance detection robustness and combine with the depth modality to augment structural cues [7]. Despite this progress, these methods remain unable to overcome the impacts of illumination attenuation in peripheral or far-field regions on laparoscopic imaging, which result in blurred landmark boundaries and reduced contrast of surrounding tissue structures (see Fig. 1). Beyond light degradation, another challenge lies in balancing positioning accuracy with structural continuity. While incorporating curve geometry modeling enhances landmark continuity [15], it may hinder precise endpoint localization due to excessive smoothing priors. Accordingly, overcoming the "localization-continuity tradeoff" is crucial for stabilizing the consistency of predicted liver landmarks.

In this work, we present A2ONet, an Attenuation-resilient Alternating Optimization Network that unifies appearance compensation and structure-aware refinement within a unified framework. To address illumination attenuation, we introduce an *illumination field compensation* (IFC) block that adaptively restores underexposed regions while preserving structural coherence. Moreover, a lightweight *frequency–orientation selective filter* (FOSF) is embedded to suppress texture-driven ambiguities and enhance salient curvilinear structures. Building upon these attenuation-resilient representations, we further design an *alternating seg–curve optimization* (ASCO) decoder that couples a segmentation branch for dense localization and a curve branch for continuous geometry. ASCO iteratively injects curve priors to regularize fragmented segmentation responses and uses refined segmentation features to guide subsequent curve updates, resulting in cascaded and mutually reinforcing optimization. The entire model is trained end-to-end with joint supervision over masks, curves, and illumination conditions. Extensive experiments on three benchmarks, L3D [19], L3D-2K [6], and P2ILF [1] demonstrate consistent improvements over state-of-the-art methods,

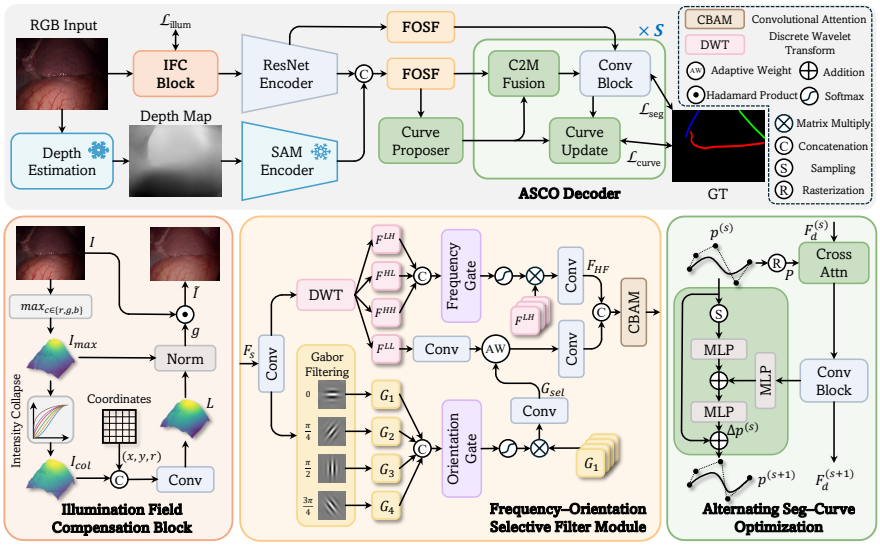


Fig. 2. Overall architecture of proposed A2ONet for liver landmark detection.

e.g., reaching 53.51% DSC and 38.32% IoU on the L3D-2K dataset, establishing a more reliable foundation for intraoperative anatomical guidance.

2 Methodology

As illustrated in Fig. 2, the proposed A2ONet is composed of four key components: an illumination field compensation (IFC) block, a multi-modal encoder, a lightweight frequency–orientation selective filter (FOSF), and a dual-branch alternating seg–curve optimization (ASCO) decoder. Given an input RGB keyframe, A2ONet first adopts the IFC block to normalize varying illumination for mitigating peripheral attenuation and then employs a ResNet [9] to capture appearance representations from the compensated RGB image. Concurrently, following the design of D2GPLand+ [6], Depth Anything V2 [25] estimates a monocular depth map from the RGB input to supply global geometric cues complementary to visual appearance [19], followed by a frozen SAM image encoder [12] to extract depth-aware geometry features. Subsequently, we concatenate RGB-D features and utilize FOSF modules to operate on the bottleneck features, emphasizing curvilinear structures while suppressing repetitive textures. We further include a curve proposer to initialize coarse parametric curves from the deepest FOSF-enhanced feature. Building on these initial curves, the ASCO decoder integrates a segmentation branch with FOSF-based skip connections for dense localization and a curve branch for explicit landmark geometry parameterization. In successive refinement stages, injected curve priors regularize segmentation responses through curve-to-mask (C2M) fusion, and the mask features in turn guide curve updates.

2.1 Illumination Field Compensation Block

To enhance detection robustness under low-light laparoscopic environment, we design an IFC block that explicitly normalizes spatially varying illumination. Given an input RGB image $I \in \mathbb{R}^{H\times W\times 3}$, we adopt a standard reflectance-illumination decomposition assumption: $I(x) = R(x) \odot L(x)$, where $L(x)$ denotes a spatial illumination field and $R(x)$ represents illumination-invariant reflectance. Our objective is to estimate a smooth low-frequency $L(x)$ and compensate illumination while preserving structural consistency. We first derive a collapsed intensity map from channel-wise maximum: $I_{max}(x) = \max_{c\in\{r,g,b\}} I_c(x)$, followed by an adaptive intensity collapse:

$$I_{\text{col}}(x) = \left(\sin\left(\frac{\pi I_{max}(x)}{2}\right) + \epsilon\right)^{\frac{1}{k}}, k > 0, \tag{1}$$

where k is a learnable temperature parameter that modulates the dynamic-range compression intensity, enabling stable illumination estimation while maintaining details. The transformed intensity is concatenated with normalized spatial coordinates (x, y) and radial distance r to estimate $L(x)$, which is then used to normalize intensity: $I_{\text{norm}}(x) = I_{max}(x)/(L(x) + \epsilon)$. To prevent over-amplification, we derive a residual gain map: $g_{\text{raw}}(x) = I_{\text{norm}}(x)/(I_{max}(x) + \epsilon)$, and apply residual blending in the gain domain: $g(x) = 1 + \alpha\big(g_{\text{raw}}(x) - 1\big)$, where $\alpha \in [0, 1]$ controls the compensation strength. Modulating the original image produces the final illumination-compensated output: $\tilde{I}(x) = I(x) \odot g(x)$ with $g(x)$ bounded to a valid range, producing a illumination-normalized RGB image that is highly invariant to low-light degradation.

2.2 Frequency–Orientation Selective Filter

To enhance the perception of curvilinear features while suppressing texture-induced ambiguities, we design the frequency–orientation selective filter (FOSF) module to operate on bottleneck features. Given an intermediate feature $F_{\text{s}} \in \mathbb{R}^{H\times W\times C}$, we first construct a single-channel guidance map $x_{\text{in}} \in \mathbb{R}^{H\times W}$ for frequency-orientation analysis. Then, the Haar wavelet transform [16] is applied to obtain four frequency sub-bands $\{F^{\text{LL}}, F^{\text{LH}}, F^{\text{HL}}, F^{\text{HH}}\}$. To retain informative high-frequency components that are highly related to curvilinear features, we perform per-pixel competition among the three high-frequency sub-bands. Concretely, a lightweight gate is adopted to process $\{|F^{\text{LH}}|, |F^{\text{HL}}|, |F^{\text{HH}}|\}$ and predict normalized weights $(w_{\text{LH}}, w_{\text{HL}}, w_{\text{HH}})$ via softmax, yielding the aggregated high-frequency response:

$$F_{\text{HF}} = w_{\text{LH}}|F^{\text{LH}}| + w_{\text{HL}}|F^{\text{HL}}| + w_{\text{HH}}|F^{\text{HH}}|, \tag{2}$$

which prevents single texture-heavy band from domination. To further emphasize elongated structures possessing coherent directions, we apply T fixed-orientation Gabor filters [17] on x_{in}, producing orientation-specific responses $\{G_j\}_{j=1}^{T}$. Then,

a lightweight gating network predicts pixel-wise orientation weights $\{\pi_j(x)\}_{j=1}^{T}$ from concatenated responses. The selected directional evidence is computed as

$$G_{\text{sel}}(x) = \sum_{j=1}^{T} \pi_j(x)\, G_j(x), \tag{3}$$

thereby adaptively selecting the dominant orientation at each spatial location. We then balance low-frequency context and oriented structure by predicting mixing weights (α, β) conditioned on $\{F^{\text{LL}}, G_{\text{sel}}\}$ to form an adaptive integration that jointly preserves coarse geometry continuity and directional coherence. The selected high-frequency response F_{HF} is concatenated subsequently and the fused features are finally refined by a lightweight attention block [23] and forwarded to the decoder for subsequent optimization.

2.3 Alternating Seg–Curve Optimization

To jointly model dense appearance cues and explicit curvilinear geometry, the alternating seg–curve optimization (ASCO) decoder utilizes two coupled branches to iteratively refine segmentation prediction and curve estimation. To be specific, we model each landmark as a Bezier curve C parameterized by M control points $\{p_m\}_{m=1}^{M}$, which ensures geometric continuity while enabling differentiable control-point updates. At the coarsest decoding stage, initial curve hypotheses are generated in an anchor-based manner, where each spatial location provides a normalized coordinates $a(x) \in (0,1)^2$. For each landmark class m, the proposer predicts confidence scores and regresses control-point offsets $\Delta_m(x)$ in logit space: $p_m(x) = \sigma(\text{logit}(a(x)) + \Delta_m(x))$, where $\text{logit}(x) = \log(x/(1-x))$ with x clamped to $[10^{-6}, 1-10^{-6}]$ for numerical stability, and $\sigma(\cdot)$ is the sigmoid function. The top-K sets of anchors are selected as initial geometric hypotheses, providing coarse structural initialization. To inject curve priors into dense prediction, the predicted curve is softly rasterized into a spatial prior map:

$$P(x) = \exp\left(-\frac{1}{2\sigma^2} \min_{t\in[0,1]} \|x - C(t)\|_2^2\right). \tag{4}$$

Through cross-attention, the rasterized prior constrains decoder features to align with the underlying curve geometry. Starting from the coarse proposals, the curve branch undergoes refinement using segmentation features across decoding stages. At stage s, decoder features $F_d^{(s)}$ are sampled along the current curve hypothesis $C^{(s)}$, yielding a curve embedding that predicts control-point updates:

$$p_k^{(s+1)} = p_k^{(s)} + \Delta p_k^{(s)}, \qquad \Delta p_k^{(s)} = \mathcal{R}^{(s)}\left(\Phi\left(F_d^{(s)}, C^{(s)}\right)\right), \tag{5}$$

where $\Phi(\cdot)$ denotes feature sampling and $\mathcal{R}^{(s)}(\cdot)$ is a stage-specific refinement head. Re-rasterizing the refined curve and re-injecting it into the segmentation branch forms a cascaded alternating loop executed S times.

2.4 Training Objectives

The network is trained end-to-end with a weighted multi-task objective:

$$\mathcal{L} = \mathcal{L}_{\text{seg}} + \lambda_{\text{curve}}\mathcal{L}_{\text{curve}} + \lambda_{\text{illum}}\mathcal{L}_{\text{illum}}. \tag{6}$$

The term $\mathcal{L}_{\text{seg}}$ supervises dense landmark masks using a combination of Dice and BCE losses to balance region overlap and pixel-wise classification. We further apply a modified center-line loss [7] on the curves to enforce structural continuity of elongated landmarks across refinement stages. Finally, to stabilize illumination compensation and avoid over-amplification, we regularize both the predicted illumination field and the gain map by penalizing spatial variations and excessive deviation from unity:

$$\mathcal{L}_{\text{illum}} = \|\nabla L\|_1 + \|g - \mathbf{1}\|_1. \tag{7}$$

3 Experiments

3.1 Datasets and Evaluation Metrics

We evaluate the proposed method on three public laparoscopic liver landmark benchmarks: L3D [19], L3D-2K [6], and P2ILF [1]. Specifically, L3D comprises 1,152 annotated keyframes, partitioned into 921 for training, 122 for validation, and 109 for testing. L3D-2K includes 2,000 keyframes, split into 1,532/230, and 238 images for training, validation, and testing, respectively. For P2ILF, which originally provides 183 annotated images with 167 allocated for training, the unavailability of the official testing set necessitates subdividing its training subset into 124 images for training and 43 for testing. To quantify region-level segmentation accuracy, we utilize the dice score coefficient (DSC) and intersection-over-union (IoU), alongside the average symmetric surface distance (ASSD) to evaluate the boundary localization of elongated landmarks.

3.2 Implementation Details

The RGB encoder employs a ResNet-34 backbone, while a frozen SAM-b image encoder extracts depth features. Geometrically, we use $T = 4$ Gabor filters with orientations $\{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$, and each landmark curve is parameterized by $M = 5$ control points refined across $S = 4$ cascaded stages with $K = 16$ initial anchors. For data preprocessing, input images are resized to 1024×1024 and augmented via random rotation and flipping. The framework is trained on a single NVIDIA GeForce RTX 3090 GPU with a batch size of 2 for 100 epochs. Optimization proceeds via AdamW with an initial learning rate of $1e-4$ and a weight decay of $1e-5$, guided by a cosine learning rate schedule to reach a minimum learning rate of $1e-6$. The weighting coefficients are set to $\lambda_{\text{curve}} = 0.2$ and $\lambda_{\text{illum}} = 0.1$.

Table 1. Quantitative comparison on L3D-2K, L3D and P2ILF datasets.

Methods	L3D-2K			L3D			P2ILF		
	DSC↑	IoU↑	ASSD↓	DSC↑	IoU↑	ASSD↓	DSC↑	IoU↑	ASSD↓
U-Net [21]	35.17	23.74	76.78	51.39	36.35	84.94	29.89	17.89	47.95
COSNet [14]	37.35	25.96	59.24	56.24	40.98	69.22	26.78	16.05	47.33
Res-UNet [24]	40.39	28.43	55.80	55.47	40.68	70.66	25.84	15.76	50.97
UNet++ [29]	38.70	27.26	58.70	57.09	41.92	74.31	34.96	21.57	42.39
HRNet [22]	41.19	29.36	56.40	58.36	43.50	70.02	33.64	21.31	44.16
DeepLabV3+ [4]	41.55	29.89	55.11	59.74	44.92	60.86	27.62	16.42	50.77
TransUNet [3]	37.54	25.85	46.75	56.81	41.44	76.16	25.49	15.22	52.14
CaseNet [26]	40.23	28.37	56.55	60.94	46.09	63.89	33.94	21.33	45.02
nnU-Net [11]	41.25	29.30	51.28	60.97	46.14	62.03	35.41	22.06	42.64
SAM-Adapter [5]	39.20	26.81	59.61	57.57	42.88	74.31	21.12	12.00	57.13
SAMed [27]	43.72	31.20	43.01	62.03	47.17	61.55	31.73	19.42	40.08
SAM-LST [2]	39.43	27.39	56.69	60.51	45.03	68.87	28.75	18.38	46.53
AutoSAM [10]	40.47	28.57	54.58	59.12	44.21	62.49	25.71	15.43	53.65
Double Res-UNet [1]	40.53	28.60	53.75	58.39	43.52	63.78	36.54	22.80	39.26
D^2GPLand [19]	49.31	35.76	41.34	63.52	48.68	59.38	40.55	25.87	38.73
D^2GPLand+ [6]	51.66	37.59	36.24	64.62	50.14	57.68	42.15	27.40	37.12
A2ONet	**53.51**	**38.32**	**30.97**	**65.31**	**50.87**	**49.89**	**43.02**	**27.95**	**33.47**

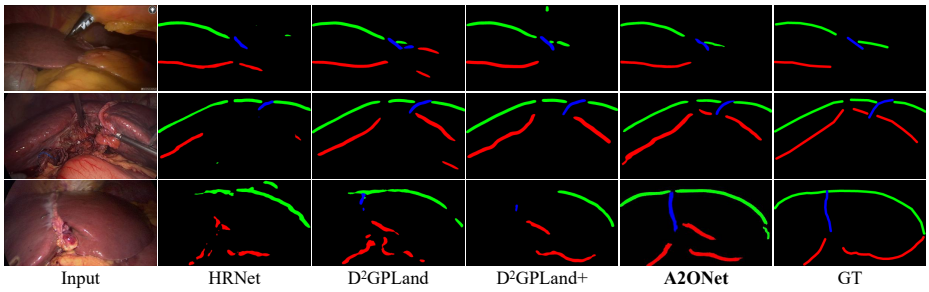


Fig. 3. Visualizations of the landmark detection results on the L3D-2K, L3D, and P2ILF datasets, arranged from top to bottom respectively.

3.3 Comparison with State-of-the-Art

We benchmark our method against representative baselines classified into three categories: (i) general-purpose segmentation models (U-Net [21], COSNet [14], Res-UNet [24], U-Net++ [29], HRNet [22], DeepLabV3+ [4], TransUNet [3], CaseNet [26], nnU-Net [11]), (ii) foundation model-based adaptation methods (SAM-Adapter [5], SAMed [27], SAM-LST [2], AutoSAM [10]), and (iii) task-specific detection approaches (D^2GPLand [19], D^2GPLand+ [6], Double Res-UNet [1]). Direct comparison excludes curve-only methods, as their evaluation protocol, *i.e.*, expanding curves to a fixed thickness for region-based assessment, precludes fair benchmarking against mask-based predictors, leaving curve-native comparisons for future work. Table 1 indicates that our method achieves the highest Dice and IoU across all datasets, establishing robust region-level delineation despite low-light and texture interference. Compared with D^2GPLand+, the strongest competing task-specific baseline, our framework improves DSC/IoU

Table 2. Ablation study of key components on the L3D-2K dataset.

Variant	IFC	$\mathcal{L}_{illum}$	FOSF	ASCO	$\mathcal{L}_{curve}$	DSC↑	IoU↑	ASSD↓
Baseline						37.69	27.13	58.76
V1	✓					38.23	27.76	56.90
V2	✓	✓				40.45	29.04	52.36
V3			✓			39.98	28.29	53.49
V4	✓	✓	✓			42.48	31.14	47.82
V5	✓	✓	✓	✓		44.45	33.10	43.33
A2ONet	✓	✓	✓	✓	✓	**45.16**	**33.76**	**40.53**

by 1.85%/0.73% on L3D-2K, 0.69%/0.73% on L3D, and 0.87%/0.55% on P2ILF. Notably, it also reduces ASSD by 5.27, 7.79, and 3.65 pixels on the three datasets, respectively, indicating more precise boundary alignment alongside enhanced landmark coverage. Further, paired significance tests confirm statistically significant improvements across datasets.

Qualitatively, Fig. 3 illustrates that existing baselines frequently generate fragmented predictions along thin curvilinear landmarks under low-light conditions. Conversely, our method yields continuous, geometrically consistent delineations with fewer false positives, demonstrating the advantages of explicit curve regularization and illumination stabilization.

3.4 Ablation Study

To verify the effectiveness of individual components in our framework, we conduct ablation experiments on the L3D-2K validation set. The baseline model consists of a ResNet encoder and a pretrained SAM image encoder, followed by a standard CNN decoder. Integrating the IFC block (V1) establishes consistent gains over this baseline, indicating the importance of explicitly normalizing low-light appearance. Further introducing $\mathcal{L}_{illum}$ (V2) yields a larger gain with 2.22% in DSC, 1.28% in IoU, and 4.54 pixels in ASSD, indicating that overexposure suppression stabilizes illumination compensation. Independent application of FOSF modules (V3) also improves performance, demonstrating effective suppression of texture-induced ambiguities. Combining IFC and $\mathcal{L}_{illum}$ with FOSF blocks (V4) produces the largest jump compared to previous variants, validating their strong complementarity. Building upon these feature-level enhancements, incorporating the ASCO decoder (V5) further improves overlap and boundary accuracy by 1.97% in DSC, 1.96% in IoU, and 4.49 pixels in ASSD, justifying the effectiveness of alternating optimization. Finally, applying $\mathcal{L}_{curve}$ yields the best overall configuration, confirming that explicit curve supervision provides extra geometric regularization beyond architectural coupling.

4 Conclusion

This study introduces A2ONet, an attenuation-resilient alternating optimization framework for reliable laparoscopic liver landmark detection. By jointly ad-

dressing illumination degradation and texture-induced ambiguities, A2ONet enhances representation robustness through illumination field compensation and frequency-orientation selective filtering. More importantly, the proposed alternating seg–curve optimization decoder explicitly reconciles dense pixel-wise localization with continuous curvilinear geometry, alleviating the inherent localization–continuity trade-off. Extensive experiments on L3D-2K, L3D, and P2ILF demonstrate consistent improvements over strong baselines, while comprehensive ablation studies confirm the complementary effectiveness of our core components. By yielding structurally consistent landmarks under challenging illumination, A2ONet provides a stable and robust prerequisite for AR-guided navigation.

Acknowledgments. This work 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 (No. SGDX20230821092359002), in part by the Research Grants Council of the Hong Kong Special Administrative Region, China, under Project T45-401/22-N, and in part by a project under the Collaborative Research with World-leading Research Groups scheme of The Hong Kong Polytechnic University (No. G-SACF).

Disclosure of Interests. The authors have no competing interests.

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. Medical image analysis **99**, 103371 (2025)
2. 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)
3. 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)
4. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV). pp. 801–818 (2018)
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: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3367–3375 (2023)
6. Cui, R., Si, W., Li, Z., Wang, K., Pei, J., Heng, P.A., Qin, J.: Depth-induced prompt learning for laparoscopic liver landmark detection. Medical Image Analysis p. 103940 (2026)
7. Cui, R., Zhang, J., Pei, J., Wang, K., Heng, P.A., Qin, J.: Topology-constrained learning for efficient laparoscopic liver landmark detection. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 585–594. Springer (2025)

8. Guo, D., Si, W., Li, Z., Pei, J., Heng, P.A.: Surgical workflow recognition and blocking effectiveness detection in laparoscopic liver resection with pringle maneuver. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 3220–3228 (2025)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
10. Hu, X., Xu, X., Shi, Y.: How to efficiently adapt large segmentation model (sam) to medical images. arXiv preprint arXiv:2306.13731 (2023)
11. Isensee, F., Wald, T., Ulrich, C., Baumgartner, M., Roy, S., Maier-Hein, K., Jaeger, P.F.: nnu-net revisited: A call for rigorous validation in 3d medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 488–498. Springer (2024)
12. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4015–4026 (2023)
13. Koo, B., Robu, M.R., Allam, M., Pfeiffer, M., Thompson, S., Gurusamy, K., Davidson, B., Speidel, S., Hawkes, D., Stoyanov, D., et al.: Automatic, global registration in laparoscopic liver surgery. International Journal of Computer Assisted Radiology and Surgery **17**(1), 167–176 (2022)
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: Medical imaging with deep learning (2023)
15. Li, Q., Liu, F., Yang, S., Shen, D., Jin, Y.: Bcrnet: Enhancing landmark detection in laparoscopic liver surgery via bezier curve refinement. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 77–87. Springer (2025)
16. Mallat, S.G.: A theory for multiresolution signal decomposition: the wavelet representation. IEEE transactions on pattern analysis and machine intelligence **11**(7), 674–693 (2002)
17. Mehrotra, R., Namuduri, K.R., Ranganathan, N.: Gabor filter-based edge detection. Pattern recognition **25**(12), 1479–1494 (1992)
18. Nguyen, K.T., Marsh, J.W., Tsung, A., Steel, J.J.L., Gamblin, T.C., Geller, D.A.: Comparative benefits of laparoscopic vs open hepatic resection: a critical appraisal. Archives of surgery **146**(3), 348–356 (2011)
19. Pei, J., Cui, R., Li, Y., Si, W., Qin, J., Heng, P.A.: Depth-driven geometric prompt learning for laparoscopic liver landmark detection. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 154–164. Springer (2024)
20. Plantefeve, R., Haouchine, N., Radoux, J.P., Cotin, S.: Automatic alignment of pre and intraoperative data using anatomical landmarks for augmented laparoscopic liver surgery. In: International Symposium on Biomedical Simulation. pp. 58–66. Springer (2014)
21. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
22. 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 transactions on pattern analysis and machine intelligence **43**(10), 3349–3364 (2020)

23. Woo, S., Park, J., Lee, J.Y., Kweon, I.S.: Cbam: Convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV). pp. 3–19 (2018)
24. Xiao, X., Lian, S., Luo, Z., Li, S.: Weighted res-unet for high-quality retina vessel segmentation. In: 2018 9th international conference on information technology in medicine and education (ITME). pp. 327–331. IEEE (2018)
25. Yang, L., Kang, B., Huang, Z., Zhao, Z., Xu, X., Feng, J., Zhao, H.: Depth anything v2. Advances in Neural Information Processing Systems **37**, 21875–21911 (2024)
26. Yu, Z., Feng, C., Liu, M.Y., Ramalingam, S.: Casenet: Deep category-aware semantic edge detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5964–5973 (2017)
27. Zhang, K., Liu, D.: Customized segment anything model for medical image segmentation. arXiv preprint arXiv:2304.13785 (2023)
28. Zhou, J., Gao, B., Wang, K., Pei, J., Heng, P.A., Qin, J.: Landmark-free preoperative-to-intraoperative registration in laparoscopic liver resection. IEEE Transactions on Medical Imaging (2025)
29. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE transactions on medical imaging **39**(6), 1856–1867 (2019)