Synergistic Bleeding Region and Point Detection in Laparoscopic Surgical Videos
⟨ Supplementary Material ⟩

Jialun Pei[1] Zhangjun Zhou[2] Diandian Guo[1] Zhixi Li[2,3]
Jing Qin[2] Bo Du[4*] Pheng-Ann Heng[1]

[1] The Chinese University of Hong Kong [2] The Hong Kong Polytechnic University
[3] Southern Medical University [4] Wuhan University

A. Overview

We summarize the supplementary material from the following aspects:

Section B: SurgBlood Dataset: More data exhibitions for SurgBlood dataset.

Section C: Exploring SAM 2 for Multi-Task Learning: Performance Comparison of SAM2 Variants and BlooDet.

Section D: Ablation Studies: We implement more ablation experiments to improve the performance of BlooDet, including 1) Ablations for different numbers of input frames. 2) Performance on different scale versions from SAM 2. 3) Ablations for frozen PWC-Net.

Section E: Pseudocode for BlooDet: We provide the pseudocode for bidirectional guidance of BlooDet.

Section F: More Comparison Experiments: We further conduct the fine-grained attribute-level experiment on SurgBlood dataset and generalization experiment on Rabbani *et al.* [11] dataset.

Section G: Additional Visual Results: We provide additional visual comparisons of BlooDet and other methods.

B. SurgBlood Dataset

B.1. More Data Visualization

The video data source in SurgBlood is from our partner hospitals, which contains 95 video clips of laparoscopic surgery from 42 patients who required cholecystectomy. To facilitate a more comprehensive understanding and analysis of SurgBlood, we divide the bleeding scenes into four categories according to specific tissues: Gallbladder, Cystic triangle, Vessel, and Gallbladder bed. Fig. 1 shows more sequential frames of bleeding scenes for different tissue locations as well as region-level and point-level annotations. As can be seen, the diverse bleeding tissues and types fully reflect the great challenges posed by bleeding detection in real laparoscopic surgery videos.

*Corresponding author. (dubo@whu.edu.cn)

C. Exploring SAM 2 for Multi-Task Learning

In Table 1, we explore various SAM 2 configurations for performing two related tasks: bleeding region detection and bleeding point localization. We first maintain the original SAM 2 setting and train it only for the region detection task. The results show that full fine-tuning performs better than decoder-only fine-tuning. Next, following classical multi-task architectures in medical image analysis [4, 9, 10], we extend SAM 2 by adding a point prediction head to enable dual-task learning. When we freeze the trained region detection network and fine-tune only the point head, the point detection accuracy remains unsatisfactory. However, jointly fine-tuning both tasks improves point detection but significantly degrades segmentation performance, suggesting that simply adding an auxiliary head leads to task interference. To mitigate this issue, we further adopt a dual-branch design with shallow feature sharing, which alleviates the performance drop in region detection but still fails to fully recover the single-task accuracy. These observations motivate us to develop BlooDet, a more effective multi-task framework that achieves superior performance on both bleeding region and point detection simultaneously.

D. Ablation Studies

D.1. Different Numbers of Input Frames

As shown in Table 2, when the number of input frames increases from 2 to 8, the IoU and PCK-2% metrics improve by 10.26% and 8.74%, respectively. This finding confirms the effectiveness of leveraging long-term temporal information. Beyond eight frames, the performance of BlooDet shows no further significant improvement. Finally, we select to input eight consecutive frames to make our method achieve the optimal performance.

D.2. Baselines for Different SAM 2 Models

We also conduct an ablation study to assess the image encoder of our framework based on different scale models of

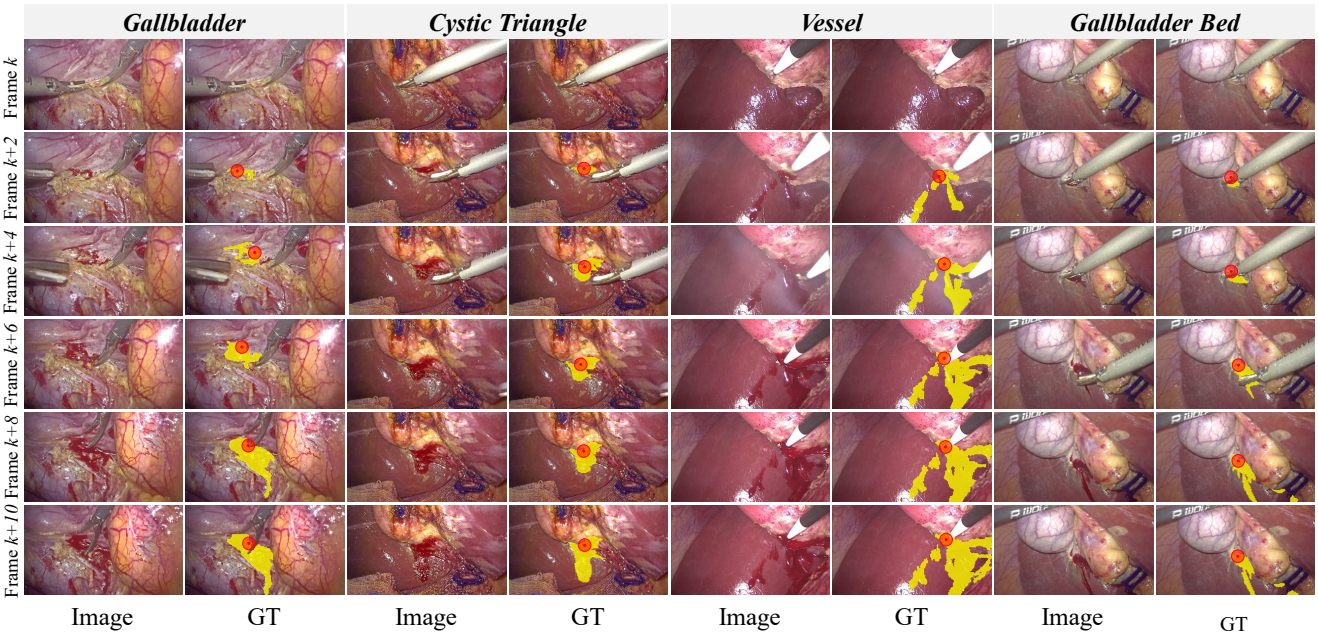

Figure 1. Examples of annotations for bleeding regions and points in SurgBlood. Zoom in for details.

Table 1. Comparison between BlooDet and different SAM2 settings in single-task and multi-task scenarios.

Setting	**Methods**	**IoU↑**	**Dice↑**	**PCK-2%↑**	**PCK-5%↑**	**PCK-10%↑**
Single-task Single-branch	SAM2 (Full-Tuning)	63.51	77.68	-	-	-
	SAM2 (Tuning decoder only)	55.59	71.46	-	-	-
Dual-task Single-branch (dual head)	SAM2' (Frozen Segment, Tuning point-head only)	63.51	77.68	3.13	19.11	47.94
	SAM2† (Full-Tuning)	50.93	67.49	12.35	41.68	71.99
Dual-task Dual-branch (shallow feature sharing)	SAM2* (Full-Tuning)	56.22	71.98	13.18	45.14	75.78
	BlooDet (Full-Tuning)	**64.88**	**78.70**	**18.62**	**55.85**	**83.69**

Table 2. Ablation for the number of input consecutive frames on BlooDet for bleeding region and point detection.

Number	IoU ↑	Dice ↑	PCK-2% ↑	PCK-5% ↑	PCK-10% ↑
2	54.62	70.65	9.88	32.95	70.68
4	58.93	74.16	8.40	38.39	72.98
8	64.88	78.70	**18.62**	**55.85**	83.69
10	**65.63**	**79.25**	13.51	54.70	**85.34**

Table 3. Performance of BlooDet based on different scale models of SAM 2 on the SurgBlood test set.

Models	IoU	Dice	PCK-2%	PCK-5%	PCK-10%	# Params
SAM 2_tiny	62.91	77.23	8.07	40.53	74.14	**49.7**M
SAM 2_small	64.10	77.69	13.58	46.56	78.13	56.8M
SAM 2_base	**64.88**	**78.70**	**18.62**	**55.85**	**83.69**	91.6M

SAM 2 on bleeding region and bleeding point detection. As illustrated in Table 3, the performance of SAM 2_tiny is noticeably worse than that of SAM 2_small and SAM 2_base. Moreover, although SAM 2_small and SAM 2_base perform similarly in detecting bleeding regions, SAM 2_base provides a clear advantage in the detection of bleeding points, *e.g.*, with a 5.04% improvement in PCK-2% score. To ensure the accuracy of bleed point detection, we ultimately decided to build our backbone based on SAM 2_base to fulfill surgeons' needs for bleeding point assisted detection.

D.3. Ablations for Frozen PWC-Net

We also conduct ablation experiments to evaluate the impact of fine-tuning PWC-Net [13] with optical flow estimation on bleeding region and point detection tasks. As shown in Table 4, although the fine-tuned model achieved comparable performance in bleeding region detection, its performance on bleeding point detection decreased. One possible reason is that in point memory modeling, the operation of the optical flow model with the reverse bleeding mask causes the network to tend toward region-level motion

Table 4. Effect of frozen PWC-Net [13] on BlooDet for bleeding region and point detection with SAM 2_tiny backbone.

Frozen	IoU ↑	Dice ↑	PCK-2% ↑	PCK-5% ↑	PCK-10% ↑	Tune Params ↓
✓	**54.62**	**70.65**	**9.88**	**32.95**	**70.68**	**49.7**M
✗	54.57	70.41	8.22	30.45	67.66	59.1M

estimation, thereby weakening the fine-grained pixel displacement estimation for accurate bleeding point localization. Moreover, fine-tuning PWC-Net resulted in approximately a 19% increase in number of tunable parameters. Thus, we prefer to keep PWC-Net frozen during training.

Algorithm 1: Synergistic Bleeding Region and Point Detection in BlooDet

Input: Video frames V, maximum number of iterations K, T
Output: Parameters $\boldsymbol{\theta}$, $\boldsymbol{\vartheta}$ for mask and point branches, and $\boldsymbol{\alpha}$ for feature encoder.
Initialize $\boldsymbol{\alpha}$ with pre-trained SAM 2;
for $k \leftarrow 1$ **to** K **do**
  for *video frame* $I_k \in V$ **do**
    $F_k \leftarrow \boldsymbol{\alpha}(I_k)$
    // **Mask Branch**
    $F^k_{mask} \leftarrow \text{MemoryModeling}(F_k, M^m_{k-1})$;
    edge_prompts $E^k_m \leftarrow \text{EdgeGenerator}(F^k_{mask})$;
    multi_prompts $\leftarrow (E^k_m,$ point_prompts $P^{k-1}_m)$;
    predict_mask $M_k \leftarrow$ MaskDecoder(multi_prompts, F^k_{mask});
    $M^m_k, M_k \leftarrow \text{UpdateMemory}(F_k, M_k)$;
    // **Point Branch**
    OpticalFlow $O_k \leftarrow \text{OpticalFlowEstimation}(I_k)$;
    $F^k_{point} \leftarrow$ MemoryModeling$(F_k, M_k, M^m_k, O_k, M^p_{k-1})$;
    predict_points $P^k_m \leftarrow \text{PointDecoder}(F^k_{point})$;
    $M^p_k \leftarrow \text{UpdateMemory}(P^k_m)$;
    // **Cross-Task Alternating Learning**
    for $t \leftarrow 1$ **to** T **do**
      $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \eta_\theta \nabla_{\boldsymbol{\theta}} \mathcal{L}_{\mathtt{m}}\Big(\boldsymbol{\theta}^{(t)}\big(\boldsymbol{\vartheta}^{(t)}\big)\Big)$;
      $\boldsymbol{\vartheta}^{(t+1)} = \boldsymbol{\vartheta}^{(t)} - \eta_\vartheta \nabla_{\boldsymbol{\vartheta}} \mathcal{L}_{\mathtt{p}}\Big(\boldsymbol{\vartheta}^{(t)}\big(\boldsymbol{\theta}^{(t+1)}\big)\Big)$;
    end
  end
  return $\boldsymbol{\theta}$, $\boldsymbol{\vartheta}$, $\boldsymbol{\alpha}$;
end
return $\boldsymbol{\theta}$, $\boldsymbol{\vartheta}$, $\boldsymbol{\alpha}$;

E. Pseudocode for BlooDet

We outline the training procedure for our synergistic bleeding region and point detection framework in Algorithm 1. Given video frames V, our model iteratively updates tuned parameters over T steps. Each frame I_k is encoded into features F_k via the shared encoder θ. Mask branch utilizes an edge generator to obtain mask prompt embedding E^k_m and combines point prompt P^k_m from point branch to generate bleeding mask and update mask memory modeling.

Table 5. Comparison on Rabbani *et al.* [11] dataset. Scores are reported as mean ± standard deviation. † and ∗ denote region- and point-level models with an additional point/region prediction head for dual-task learning.

Methods	Volumes	IoU ↑	Dice ↑	Params ↓
Swin-UNet† [1]	ECCV'22	24.96 ± 0.14	39.95 ± 0.23	27.2M
SAM† [7]	ICCV'23	33.58 ± 0.11	49.61 ± 0.19	93.9M
MemSAM† [5]	CVPR'24	34.25 ± 0.21	50.81 ± 0.25	133.5M
STDDNet† [2]	ICCV'25	26.92 ± 0.17	42.42 ± 0.24	38.2M
SAM 2† [12]	ICLR'25	35.06 ± 0.16	52.08 ± 0.22	81.0M
HRNet∗ [14]	CVPR'19	28.98 ± 0.13	44.93 ± 0.18	63.6M
SimCC∗ [8]	ECCV'22	34.83 ± 0.20	51.75 ± 0.25	66.3M
GTPT∗ [15]	ECCV'24	23.84 ± 0.12	38.51 ± 0.23	16.7M
D-CeLR∗ [3]	ECCV'24	27.15 ± 0.14	42.71 ± 0.19	53.4M
CalibratedSL∗ [6]	TMI'25	31.66 ± 0.17	48.09 ± 0.21	6.6M
PAINet [4]	MICCAI'23	33.71 ± 0.11	49.93 ± 0.18	13.6M
PitSurgRT [9]	IJCARS'24	31.41 ± 0.15	47.81 ± 0.20	67.3M
ConsisTNet [10]	IJCARS'25	36.40 ± 0.16	53.37 ± 0.19	143.5M
BlooDet (Ours)	-	**39.31** ± 0.12	**56.43** ± 0.14	91.6M

Point branch uses the prediction mask M and mask memory feature M^m from the mask branch to predict the bleeding point. After each forward process, losses for mask $\mathcal{L}_{mask}$, point $\mathcal{L}_{point}$, edge $\mathcal{L}_{edge}$, and confidence score $\mathcal{L}_{score}$ are computed to update model parameters.

F. More Comparison Experiments

F.1. Generalization Evaluation on Rabbani

To further evaluate the generalization ability of our method, we also conduct zero-shot testing on the public dataset released by Rabbani *et al.* [11] using pre-trained models from SurgBlood. This dataset consists of 751 frames extracted from gynecologic laparoscopic surgery videos, each containing active bleeding regions rather than continuous video clips. To ensure a fair comparison with video-based models, we replicated each static frame into an 8-frame clip during evaluation. As shown in Table 5, our method demonstrates superior generalization performance on this dataset compared to existing approaches.

F.2. Fine-grained evaluation on SurgBlood

According to the bleeding types of surgical video clips in the SurgBlood dataset, we also provide the performance of our model and other models on four bleeding types by tissue location: gallbladder, cystic triangle, vessel, and gallbladder bed. As shown in Table 6, BlooDet has a clear advantage in most of the attributes and ranks ahead in terms of both bleeding region and point metrics, *e.g.*, achieving 72.34 and 36.00 scores on IoU and PCK-2% metrics. The fine-grained results further illustrate that our framework exhibits robust performance across various types of bleeding.

Table 6. Results of SurgBlood and other methods for four types on SurgBlood test set. † and ∗ denote region- and point-level models with an additional point/region prediction head for dual-task learning.

Types	Methods	Volumes	Bleeding Region Metrics		Bleeding Point Metrics			Params ↓
			IoU ↑	Dice ↑	PCK-2% ↑	PCK-5% ↑	PCK-10% ↑	
Gallbladder	**Swin-UNet**† [1]	ECCV'22	47.51	64.42	0.00	5.71	22.86	27.2M
	SAM† [7]	ICCV'23	44.70	61.78	2.86	14.29	52.86	93.7M
	MemSAM† [5]	CVPR'24	56.15	71.92	2.86	27.86	64.29	133.3M
	STDDNet† [2]	ICCV'25	26.83	42.31	0.71	1.43	20.00	38.2M
	SAM 2† [12]	ICLR'25	61.91	76.47	20.00	57.14	88.57	81.0M
	HRNet∗ [14]	CVPR'19	53.53	69.73	0.71	11.43	35.71	63.6M
	SimCC∗ [8]	ECCV'22	49.45	66.17	6.43	30.00	54.29	66.3M
	GTPT∗ [15]	ECCV'24	46.53	63.51	0.71	2.86	16.43	16.7M
	D-CeLR∗ [3]	ECCV'24	50.10	66.76	0.71	15.71	56.43	53.4M
	CalibratedSL∗ [6]	TMI'25	39.34	56.46	2.14	12.14	45.71	6.6M
	PAINet [4]	MICCAI'23	56.41	72.13	7.85	25.71	70.71	13.6M
	PitSurgRT [9]	IJCARS'24	47.39	64.30	0.71	3.57	24.28	67.3M
	ConsisTNet [10]	IJCARS'25	43.36	60.49	12.86	45.00	85.00	143.5M
	SurgBlood (Ours)	-	**62.46**	**76.90**	**25.71**	**67.85**	**91.42**	91.6M
Cystic triangle	**Swin-UNet**† [1]	ECCV'22	61.72	76.33	4.00	28.00	70.00	27.2M
	SAM† [7]	ICCV'23	55.23	71.16	4.00	23.00	63.00	93.7M
	MemSAM† [5]	CVPR'24	61.29	76.00	2.00	20.00	63.00	133.3M
	STDDNet† [2]	ICCV'25	45.00	62.07	0.00	19.00	65.00	38.2M
	SAM 2† [12]	ICLR'25	70.69	82.83	16.00	75.00	91.00	81.0M
	HRNet∗ [14]	CVPR'19	66.44	79.84	3.00	17.00	50.00	63.6M
	SimCC∗ [8]	ECCV'22	69.61	82.08	2.00	10.00	35.00	66.3M
	GTPT∗ [15]	ECCV'24	66.36	79.78	3.00	53.00	71.00	16.7M
	D-CeLR∗ [3]	ECCV'24	60.49	75.38	3.00	31.00	75.00	53.4M
	CalibratedSL∗ [6]	TMI'25	61.90	76.46	6.00	31.00	71.00	6.6M
	PAINet [4]	MICCAI'23	66.43	79.83	0.00	2.00	16.00	13.6M
	PitSurgRT [9]	IJCARS'24	64.32	78.28	4.00	14.00	55.00	67.3M
	ConsisTNet [10]	IJCARS'25	60.81	75.63	6.00	49.00	86.00	143.5M
	SurgBlood (Ours)	-	**72.34**	**83.95**	**36.00**	**80.00**	**95.00**	91.6M
Vessel	**Swin-UNet**† [1]	ECCV'22	23.09	37.51	6.67	26.67	43.33	27.2M
	SAM† [7]	ICCV'23	27.95	43.68	6.67	33.33	76.67	93.7M
	MemSAM† [5]	CVPR'24	36.88	53.88	6.67	46.67	73.33	133.3M
	STDDNet† [2]	ICCV'25	38.86	55.97	3.33	26.67	60.00	38.2M
	SAM 2† [12]	ICLR'25	48.37	65.20	26.66	56.66	70.00	81.0M
	HRNet∗ [14]	CVPR'19	33.76	50.47	0.00	10.00	36.67	63.6M
	SimCC∗ [8]	ECCV'22	37.81	54.88	0.00	23.33	53.33	66.3M
	GTPT∗ [15]	ECCV'24	15.93	27.49	0.00	3.33	20.00	16.7M
	D-CeLR∗ [3]	ECCV'24	36.64	53.63	3.33	16.67	80.00	53.4M
	CalibratedSL∗ [6]	TMI'25	25.14	40.18	0.00	0.00	6.67	6.6M
	PAINet [4]	MICCAI'23	28.45	44.29	3.33	26.66	76.66	13.6M
	PitSurgRT [9]	IJCARS'24	26.98	42.50	3.33	30.00	60.00	67.3M
	ConsisTNet [10]	IJCARS'25	37.66	54.72	13.33	60.00	80.00	143.5M
	SurgBlood (Ours)	-	**55.62**	**71.48**	**33.33**	**63.33**	**80.00**	91.6M
Gallbladder bed	**Swin-UNet**† [1]	ECCV'22	28.64	44.53	1.19	8.01	32.94	27.2M
	SAM† [7]	ICCV'23	39.81	56.94	2.08	20.47	52.23	93.7M
	MemSAM† [5]	CVPR'24	51.68	68.14	7.12	35.61	64.99	133.3M
	STDDNet† [2]	ICCV'25	52.25	68.63	2.08	16.02	49.55	38.2M
	SAM 2† [12]	ICLR'25	46.68	63.65	6.82	23.14	59.64	81.0M
	HRNet∗ [14]	CVPR'19	45.55	62.59	2.37	18.10	52.23	63.6M
	SimCC∗ [8]	ECCV'22	48.31	65.15	3.26	13.65	50.15	66.3M
	GTPT∗ [15]	ECCV'24	36.41	53.38	4.75	11.57	39.76	16.7M
	D-CeLR∗ [3]	ECCV'24	50.39	67.02	3.86	26.41	62.31	53.4M
	CalibratedSL∗ [6]	TMI'25	31.13	47.48	0.89	5.93	19.29	6.6M
	PAINet [4]	MICCAI'23	39.47	56.61	0.89	14.24	46.29	13.6M
	PitSurgRT [9]	IJCARS'24	22.67	36.97	2.60	16.61	43.32	67.3M
	ConsisTNet [10]	IJCARS'25	35.13	52.00	4.45	20.47	54.60	143.5M
	SurgBlood (Ours)	-	**64.22**	**78.21**	**8.30**	**43.32**	**77.44**	91.6M

G. Additional Visual Results

We exhibit more visualization results of our framework and other task-related methods, *i.e.*, PAINet [4], PitSurgRT [9], and SAM 2† [12], on the SurgBlood test set. As shown in Fig. 2 and Fig. 3, BlooDet provides better performance compared to other competitors and has a high degree of consistency with ground-truth for bleeding region and bleeding point detection. This further shows the robustness and accuracy of our method.

References

[1] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *ECCV*, 2022. 3, 4

[2] Guilian Chen, Huisi Wu, and Jing Qin. Stddnet: Harnessing mamba for video polyp segmentation via spatial-aligned temporal modeling and discriminative dynamic representation learning. In *IEEE ICCV*, pages 21364–21373, 2025. 3, 4

[3] Chao Dai, Yang Wang, Chaolin Huang, Jiakai Zhou, Qilin Xu, and Minpeng Xu. A cephalometric landmark regression method based on dual-encoder for high-resolution x-ray image. In *ECCV*, pages 93–109, 2024. 3, 4

[4] Adrito Das, Danyal Z Khan, Simon C Williams, John G Hanrahan, Anouk Borg, Neil L Dorward, Sophia Bano, Hani J Marcus, and Danail Stoyanov. A multi-task network for anatomy identification in endoscopic pituitary surgery. In *MICCAI*, pages 472–482, 2023. 1, 3, 4, 5

[5] Xiaolong Deng, Huisi Wu, Runhao Zeng, and Jing Qin. Memsam: taming segment anything model for echocardiography video segmentation. In *IEEE CVPR*, pages 9622–9631, 2024. 3, 4

[6] Yong Feng, Jinzhu Yang, Lingzhi Tang, Song Sun, and Yonghuai Wang. Uncertainty quantification and quality control for heatmap-based landmark detection models. *IEEE TMI*, 2025. 3, 4

[7] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *IEEE ICCV*, pages 4015–4026, 2023. 3, 4

[8] Yanjie Li, Sen Yang, Peidong Liu, Shoukui Zhang, Yunxiao Wang, Zhicheng Wang, Wankou Yang, and Shu-Tao Xia. Simcc: A simple coordinate classification perspective for human pose estimation. In *ECCV*, pages 89–106, 2022. 3, 4

[9] Zhehua Mao, Adrito Das, Mobarakol Islam, Danyal Z Khan, Simon C Williams, John G Hanrahan, Anouk Borg, Neil L Dorward, Matthew J Clarkson, Danail Stoyanov, et al. Pitsurgrt: real-time localization of critical anatomical structures in endoscopic pituitary surgery. *IJCARS*, pages 1–8, 2024. 1, 3, 4, 5

[10] Zhehua Mao, Adrito Das, Danyal Z Khan, Simon C Williams, John G Hanrahan, Danail Stoyanov, Hani J Marcus, and Sophia Bano. Consistnet: a spatio-temporal approach for consistent anatomical localization in endoscopic pituitary surgery. *IJCARS*, pages 1–10, 2025. 1, 3, 4

[11] Navid Rabbani, Callyane Sève-d'Erceville, Nicolas Bourdel, and Adrien Bartoli. Video-based computer-aided laparoscopic bleeding management: a space-time memory neural network with positional encoding and adversarial domain adaptation. In *MIDL*, pages 961–974, 2022. 1, 3

[12] Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, et al. Sam 2: Segment anything in images and videos. In *ICLR*, 2025. 3, 4, 5

[13] Deqing Sun, Xiaodong Yang, Ming-Yu Liu, and Jan Kautz. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In *IEEE CVPR*, pages 8934–8943, 2018. 2, 3

[14] Ke Sun, Bin Xiao, Dong Liu, and Jingdong Wang. Deep high-resolution representation learning for human pose estimation. In *IEEE CVPR*, pages 5693–5703, 2019. 3, 4

[15] Haonan Wang, Jie Liu, Jie Tang, Gangshan Wu, Bo Xu, Yanbing Chou, and Yong Wang. Gtpt: Group-based token pruning transformer for efficient human pose estimation. In *ECCV*, pages 213–230, 2024. 3, 4

Figure 2. Additional visual results of bleeding region and point detection on SurgBlood test set.

Figure 3. Additional visual results of bleeding region and point detection on SurgBlood test set.