B74 (Fri-AM)

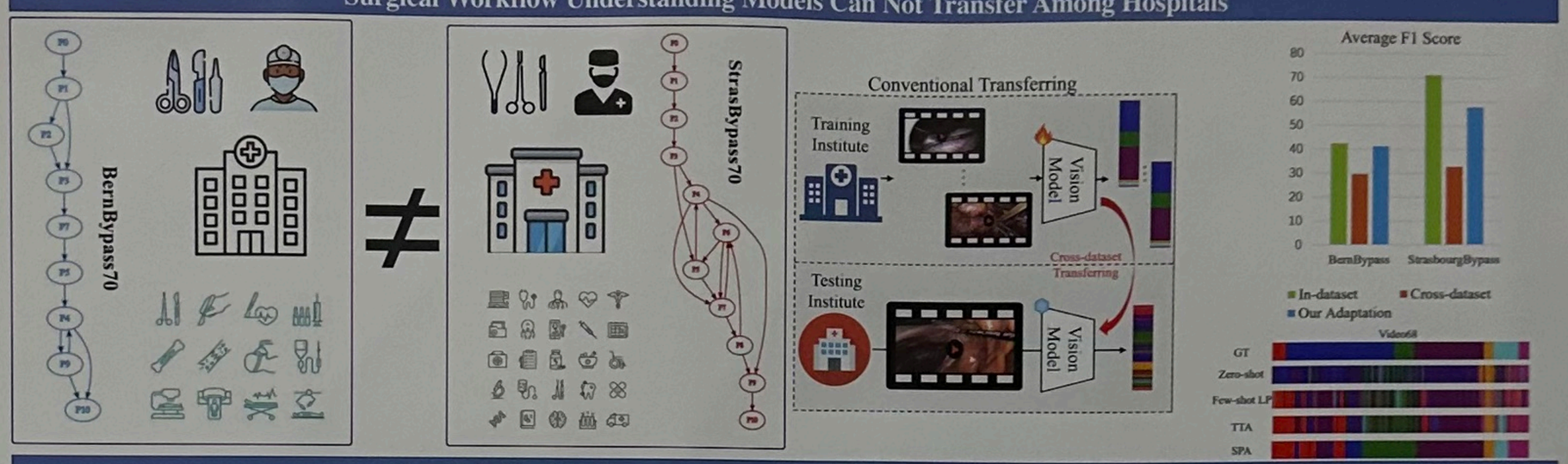
Recognizing Surgical Phases Anywhere (**SPA**)
Few-Shot Test-time Adaptation and Task-graph Guided Refinement

Kun Yuan[13], Tingxuan Chen[3], Shi Li[1], Joel L. Lavanchy[4], Christian Heiliger[5], Ege Özsoy[3], Yiming Huang[6], Long Bai[6], Nassir Navab[3], Vinkle Srivastav[12], Hongliang Ren[6], Nicolas Padoy[12]

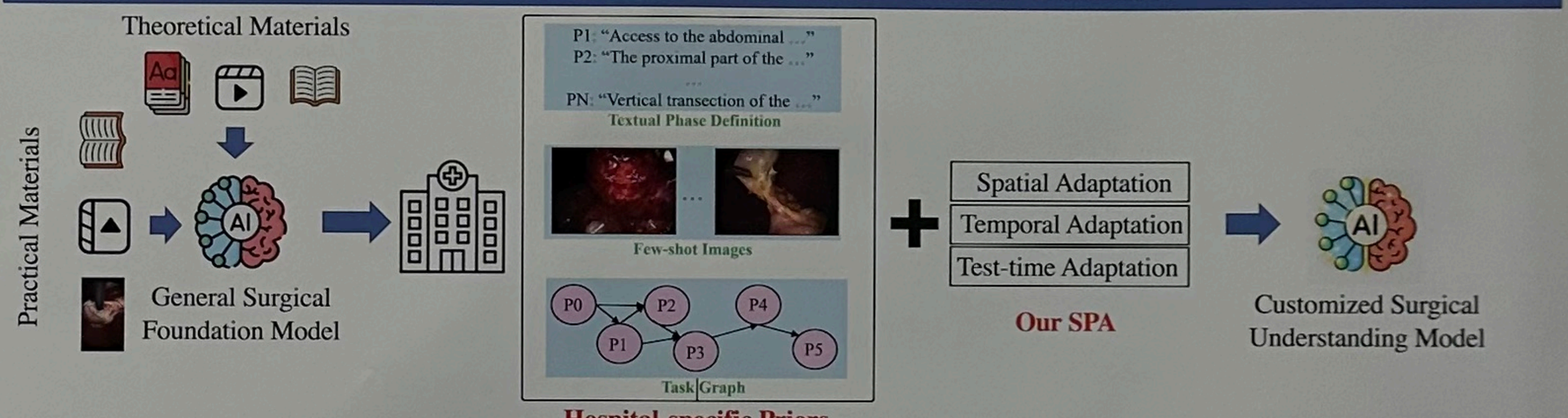
[1] University of Strasbourg, CNRS, INSERM, ICube, UMR7357, Strasbourg, France
[2] IHU Strasbourg, Strasbourg, France
[3] CAMP, Technische Universität München, Munich, Germany
[4] University Digestive Health Care Center – Clarunis, 4002 Basel, Switzerland
[5] Ludwig Maximilian University of Munich, Munich, Germany
[6] Chinese University of Hong Kong, Hong Kong SAR, China

MICCAI2025 Daejeon REPUBLIC OF KOREA

Surgical Workflow Understanding Models Can Not Transfer Among Hospitals

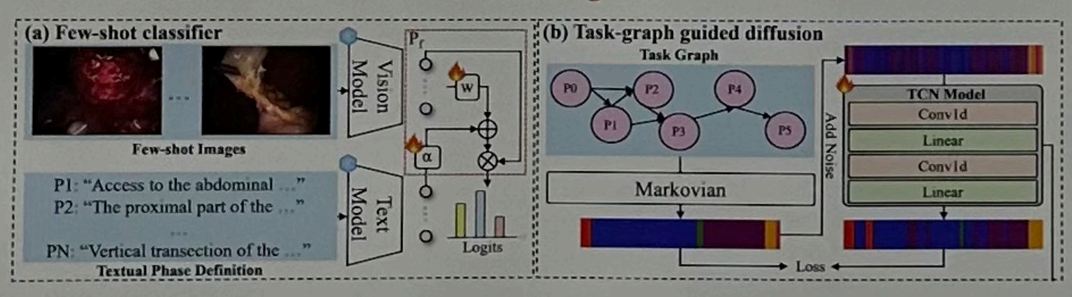


Hospitals Need to Customize Foundation Model to Their Own Domain

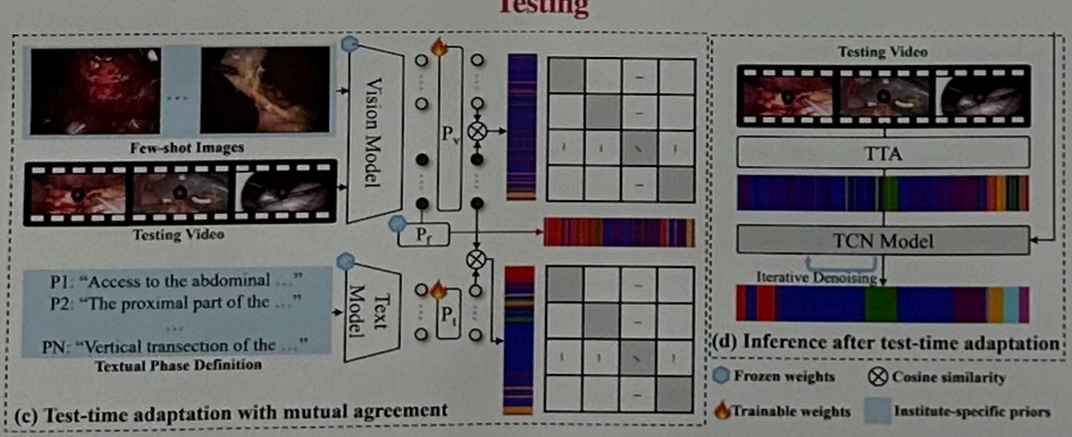


Handle Domain Gaps with Spatio-temporal Adaptations

Training



Testing



Surgical Phase Recognition

Methods	Shots	Cholec80	StrasBypass	BernBypass	Autolaparo
CLIP [15]	0	13.1	5.5	4.1	18.3
PeskaVLP [26]	0	34.2	28.6	22.6	28.6
RN50 Full	Full	80.3	71.1	42.4	NA
Tip-Adapter-F [28]	1	29.29±0.53	14.88±0.03	14.29±0.00	26.58±0.00
Linear probe (LP) [17]	1	26.19±6.75	24.70±3.64	14.95±1.35	33.90±5.19
LP+text [17]	1	32.21±1.88	23.53±3.59	14.12±2.36	32.80±5.81
SPA	1	44.53±5.77	30.68±5.24	24.31±4.29	51.48±6.26
Tip-Adapter-F [28]	16	40.75±2.15	34.29±2.12	24.54±1.97	37.51±1.31
Linear probe (LP) [17]	16	39.06±1.04	38.48±1.88	27.36±0.45	42.60±2.78
LP+text [17]	16	38.24±2.33	35.42±1.35	25.01±0.50	39.43±4.83
SPA	16	55.05±4.80	52.39±1.88	38.73±1.58	58.91±4.72
Tip-Adapter-F [28]	32	42.05±1.04	35.83±4.08	26.98±1.68	39.75±2.11
Linear probe (LP) [17]	32	38.37±1.37	42.79±0.61	29.74±0.85	44.72±1.25
LP+text [17]	32	38.36±1.81	40.35±1.25	26.01±0.79	39.40±3.35
SPA	32	55.69±3.99	55.80±1.32	43.54±1.96	60.36±3.14

Conclusion

Surgical Phase Anywhere (SPA) offers a lightweight, data-efficient framework for cross-institution surgical phase recognition. By combining few-shot spatial adaptation, task-graph-guided temporal modeling, and dynamic test-time adaptation, SPA achieves robust generalization with minimal annotation, enabling scalable and reliable deployment across diverse surgical environments.

References

[1] Yuan, K., Srivastav, V., Yu, T., Lavanchy, J.L., Marescaux, J., Mascagni, P., Navab, N., Padoy, N. Learning Multi-modal Representations by Watching Hundreds of Surgical Video Lectures. Medical Image Analysis, 2025.
[2] Radford, A., et al. Learning transferable visual models from natural language supervision.
[3] Few-shot Adaptation of Medical Vision-Language Models, MICCAI 2024.
[4] Zhang, R., et al. Tip-Adapter: Training-free CLIP-adapter for better vision-language modeling.
[5] Yuan, K., Srivastav, V., Navab, N., Padoy, N. HECVL: Hierarchical video-language pretraining for zero-shot surgical phase recognition. MICCAI, 2024.
[6] Yuan, K., Srivastav, V., Navab, N., Padoy, N., et al. Procedure-aware surgical video-language pretraining with hierarchical knowledge augmentation. arXiv:2410.00263, 2024.
[7] Lavanchy, J.L., et al. Challenges in multi-centric generalization: Phase and step recognition in Roux-en-Y gastric bypass surgery.