A86 (Wed-PM)

MICCAI2025
29th International Conference on Medical Image Computing and Computer Assisted Intervention
23-27 September 2025
Daejeon Convention Center
Daejeon REPUBLIC OF KOREA

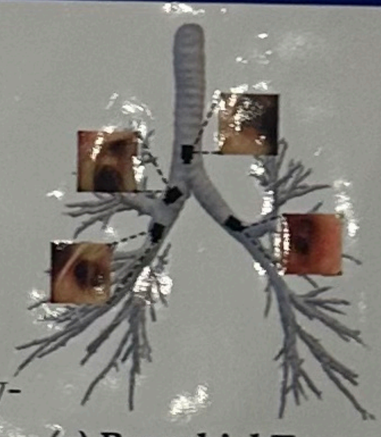
Structure-Aware Cross-Modal Prompt Tuning for Autonomous Bronchoscopic Navigation

Hao Fang[1,2], *Zhuo Zeng*[3], *Jianwei Yang*[2], *Wenkang Fan*[2], *Xiongbiao Luo*[1,2,4,*]
[1]National Institute for Data Science in Health and Medicine, Xiamen University, Xiamen 361102, China
[2]Department of Computer Science and Engineering, Xiamen University, Xiamen 361102, China
[3]Xiamen University Tan Kah Kee College, Zhangzhou 363105, China
[4]Discipline of Intelligent Instrument and Equipment, Xiamen University, Xiamen 361102, China

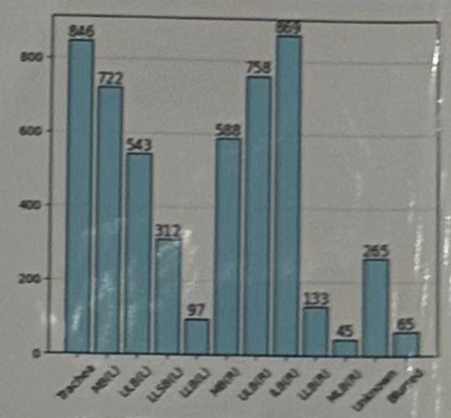
Paper ID: 3311

Introduction

- Subtle morphological variations among bronchial bifurcations may mislead surgeons, posing risks to precise positional recognition during bronchoscopy.
- Conventional recognition models rely on a fixed number of predefined classes, limiting their ability to capture numerous bronchial bifurcation variants.
- Cross-modal prompt tuning methods often fail to achieve effective feature disentanglement and to extract fine-grained low-level cues, such as textures and edge details.



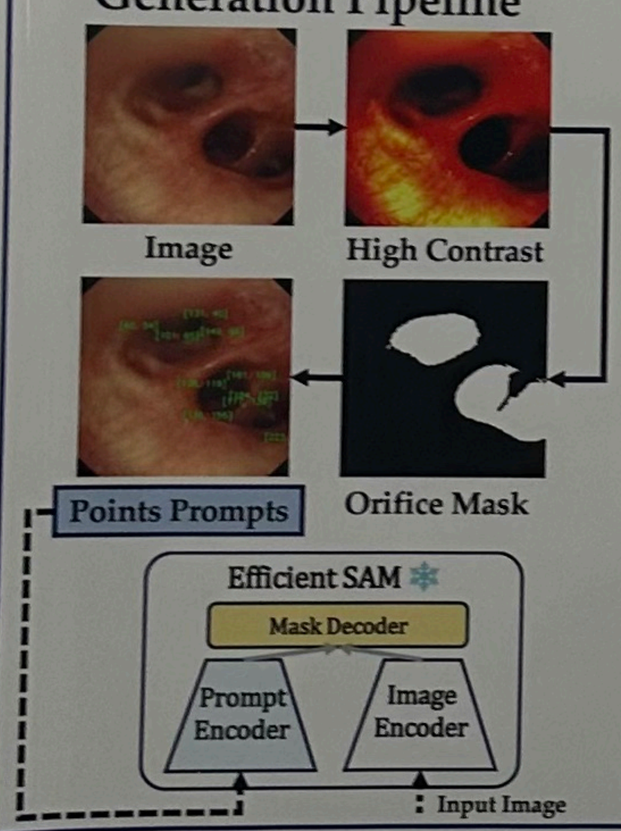
(a) Bronchial Tree



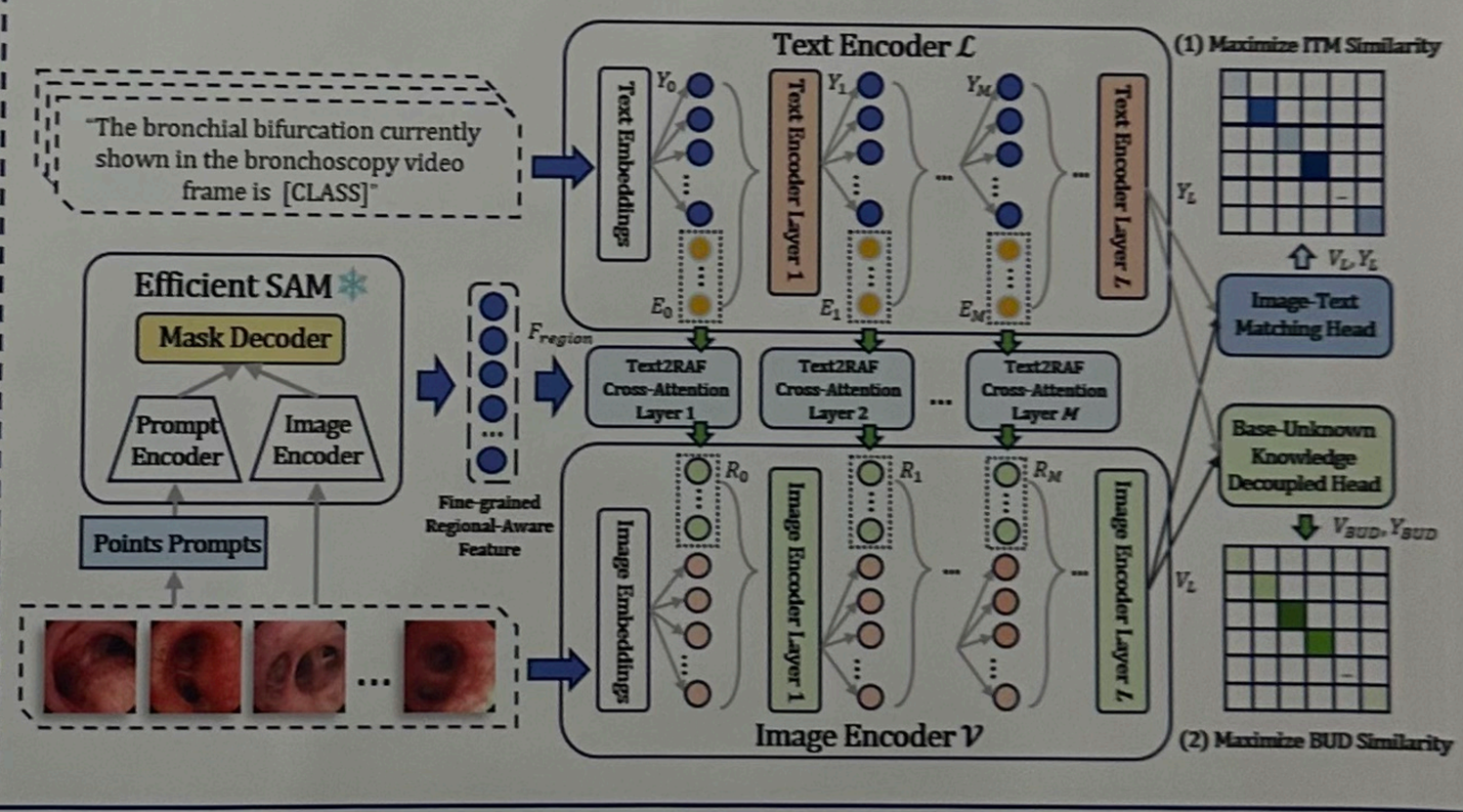
(b) Data Distribution

Methods

Automatic Point Prompts Generation Pipeline



Structure-Aware Cross-Modal Prompt Tuning Framework



Results

Recognition Performance Comparison

Methods	Accuracy	F1-score	Recall	IDR	Extra Params
Vanilla CLIP	74.94	55.31	56.56	78.00	0M
CLIP-Adapter	79.40	62.52	63.08	76.40	0.52M
CoOp	86.28	75.38	75.86	79.20	0.002M
VPT	87.10	76.36	77.06	82.60	0.074M
MaPLe	87.20	75.47	76.42	81.80	3.56M
PromptSRC	86.43	74.65	74.40	80.80	0.046M
DePT	87.35	75.52	76.11	83.10	3.57M
SCPT(Ours)	**88.94**	**79.57**	**80.71**	**87.00**	**3.28M**

t-NSE Visualization (SCPT)

Conclusion

- With fine-grained visual cues, base-unknown decoupled features in the latent space, and cross-modal prompt tuning, SCPT achieves superior performance in bronchial bifurcation recognition.
- The effectiveness of SCPT demonstrates the potential of vision-language foundation models for addressing other fine-grained and open-set medical image recognition tasks.

This work was supported in part by the National Natural Science Foundation of China under Grant 82272133, in part by the High-Quality Development Science and Technology Major Project of Xiamen Health Commission under Grant 2024GZL-ZD03, and in part by Ningbo 2035 Key Research and Development Program under Grant 2024Z127.